

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :

C12Q 1/68, C12N 7/00, A01N 43/04,
A61K 31/70, C07K 3/00, 13/00, 15/00,
17/00, C07H 17/00

A1

(11) International Publication Number:

WO 94/14980

(43) International Publication Date:

7 July 1994 (07.07.94)

(21) International Application Number: PCT/US93/12388

(22) International Filing Date: 20 December 1993 (20.12.93)

(30) Priority Data:

996,783

23 December 1992 (23.12.92)

US

123,936

17 September 1993 (17.09.93)

US

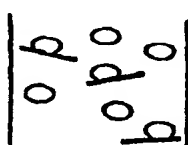
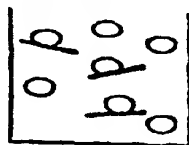
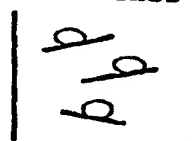
(71) Applicant: GENELABS TECHNOLOGIES, INC. [US/US];
505 Penobscot Drive, Redwood City, CA 94063 (US).(72) Inventors: EDWARDS, Cynthia, A.; 2021 Oakley Avenue,
Menlo Park, CA 94025 (US). CANTOR, Charles, R.; 640
Panoramic Way, Berkeley, CA 94707 (US). ANDREWS,
Beth, M.; 24 Franklin Street, Watertown, MA 02172 (US).
TURIN, Lisa, M.; P.O. Box 5691, Redwood City, CA 94063
(US). FRY, Kirk, E.; 2604 Ross Road, Palo Alto, CA 94303
(US).(74) Agent: FABIAN, Gary, R.; Dehlinger & Associates, P.O. Box
60850, Palo Alto, CA 94306-0850 (US).(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ,
DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, LV,
MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE,
SK, UA, UZ, VN, European patent (AT, BE, CH, DE, DK,
ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI
patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE,
SN, TD, TG).

Published

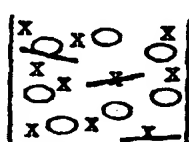
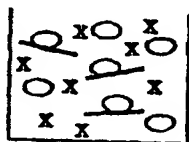
With international search report.

(54) Title: SEQUENCE-DIRECTED DNA-BINDING MOLECULES COMPOSITIONS AND METHODS

CONTROL

CAPTURE
"UNBOUND" DNACAPTURE
DNA: PROTEIN
COMPLEXES

+ INHIBITOR



LIGHT!

INCREASED
SIGNALDECREASED
SIGNAL

(57) Abstract

The present invention defines a DNA protein-binding assay useful for screening libraries of synthetic or biological compounds for their ability to bind DNA test sequences. The assay is versatile in that any number of test sequences can be tested by placing the test sequence adjacent to a defined protein-binding screening sequence. Binding of molecules to these test sequence changes the binding characteristics of the protein molecule to its cognate binding sequence. When such a molecule binds the test sequence the equilibrium of the DNA:protein complexes is disturbed, generating changes in the concentration of free DNA probe. Numerous exemplary target test sequences (SEQ ID NO:1 to SEQ ID NO:600) are set forth. The assay of the present invention is also useful to characterize the preferred binding sequences of any selected DNA-binding molecule.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

**SEQUENCE-DIRECTED DNA-BINDING MOLECULES
COMPOSITIONS AND METHODS**

5

Field of the Invention

10 The present invention relates to methods, systems,
and kits useful for the identification of molecules
that specifically bind to defined nucleic acid se-
quences. Also described are methods for designing
molecules having the ability to bind defined nucleic
15 acid sequences and compositions thereof.

Table of Contents:

	Page No.:
20 References	5
Background of the Invention	11

	Summary of the Invention	15
	Brief Description of the Figures	23
	Detailed Description of the Invention	27
	I. Definitions	27
5	II. The Assay	30
	A. General Considerations	32
	B. Choosing and Testing an Appropriate DNA-Binding Protein	34
10	1. Criteria for Choosing an Appropriate DNA-Binding Protein	35
	2. Testing DNA:Protein Interactions for Use in the Assay	36
15	a.) Other DNA: Protein Interactions Useful in the Method of the Present Invention	36
20	b.) The Use of UL9 Proteins in the Practice of the Present Invention	37
	C. Preparation of Full Length UL9 and UL9-COOH Polypeptides	38
25	D. Functional Binding and Rate of Dissociation	40
	E. <i>oriS</i> Flanking Sequence Variation	41
	F. Small Molecules as Sequence-Specific Competitive Inhibitors	43
30	G. Theoretical Considerations on the Concentration of Assay Components	49
	H. The Use of the Assay under Conditions of Very High Protein Concentration	54
35	III. Amplification-Based Selection Technique to Determine the Sequence Preferences of DNA-Binding Molecules	56
	A. Design of Test Oligonucleotides	56
40	B. Applying the Assay to the Mixed Pools of Test Oligonucleotides	57
	C. Amplification	58

	D.	Sequencing of Amplified DNA	60
	IV.	Modifications of Test Oligonucleotides and other Useful DNA:Protein Interactions	60
	V.	Capture/Detection Systems	62
5	A.	Capture of Unbound DNA	64
	1.	Modification of the Protein Recognition Sequence with Biotin	65
10	2.	Capture of Site-Specific Biotinylated Oligonucleotides	67
	B.	Capture of DNA:Protein Complexes	69
	C.	Detection Systems	69
	1.	Radioactive Labeling	70
15	2.	Chemiluminescent Detection	71
	D.	Alternative Methods for Detecting Molecules that Increase the Affinity of the DNA-Binding Protein for its Cognate Site	72
20	VI.	Utility	73
	A.	The Usefulness of Sequence-Specific DNA-Binding Molecules	73
	B.	General Applications of the Assay	75
25	1.	Mass-Screening of Libraries for the Presence of Sequence-Specific DNA-Binding Molecules	76
	2.	Directed Screening	78
30	3.	Molecules Derived from Known DNA-binding Molecules	79
	4.	Secondary Assays	80
	a.)	Confirmatory Studies	80
35	b.)	Secondary Studies to Elucidate Binding Characteristics	86
40	c.)	Restriction Endonucleases as Indicator Proteins in the Assay	88

	5.	Generation of Binding Data and Refinement of Molecular Modeling Systems	90
5	6.	The Design of New DNA-Binding Heteropolymers Comprised of Subunits Directed to Different DNA Sequences	91
10	C.	Sequences Targeted by the Assay	92
	1.	Medically Significant Target Sequences	94
	a.)	Table I: Pathogens	95
15	b.)	Table II: Non-infectious Diseases	108
	c.)	Table III: Human Genes with Promoter Regions that are Potential Targets for DNA-Binding Molecules	114
20	d.)	Table IV: Medically Significant DNA-Binding Sequences	128
	2.	Defined Sets of Test Sequences	132
25	3.	Theoretical Considerations in Choosing Biological Target Sites: Specificity and Toxicity	133
30	4.	Further Considerations in Choosing Target Sites: Finding Eukaryotic Promoters	133
35	5.	Further Considerations in Choosing Alternative Small-Molecule Binding Sites	136
40	6.	Further Considerations in Choosing Target Sites: Prokaryotes and Viruses	139
	D.	Using Test Matrices and Pattern Matching for the Analysis of Data	143

	E.	Applications for the Determination of the Sequence Specificity of DNA-Binding Drugs	146
5	1.	Multimerization of Sequence-Preferential or Sequence-Specific DNA-Binding Molecules Identified in the Assay	147
10	2.	Sequence-Specific DNA-Binding Molecules Identified in the Assay as Facilitators of Triplex Formation	152
	F.	Other Applications	155
15		Materials and Methods	156
		Examples	156
		Sequence Listing	211
		Claims	534
		Abstract	545
20		References	
		Ambinder, R.F., et al., <i>J. Virol.</i> <u>65</u> :1466-1478 (1991).	
		Angel, P., et al., <i>Nature</i> <u>332</u> :166 (1988).	
25		Ausubel, F. M., et al., <u>Current Protocols in Molecular Biology</u> , John Wiley and Sons, Inc., Media PA.	
		Baguley, B.C., <i>Mol. Cell. Bioch.</i> <u>43</u> :167-181 (1982).	
30		Banerji, S.S., et al., <i>Mol. Cell Biol.</i> <u>11</u> :4074-4087 (1991).	
		Beal, P.A., et al., <i>Science</i> <u>251</u> :1360-1363 (1991).	
		Becker, Y., et al., <i>Isr. J. Med. Sci.</i> <u>8</u> :1225 (1972).	
		Bialer, M., et al., <i>J. Med. Chem.</i> <u>23</u> :1144 (1980).	
35		Bialer, M., et al., <i>J. Pharm. Sci.</i> <u>70</u> :822 (1981).	
		Birg, F., et al., <i>Nucl. Acids Res.</i> <u>18</u> :2901-2908 (1990).	
		Bohmann, D., et al., <i>Science</i> <u>238</u> :1386 (1987).	

- Bos, T.J., et al., *Cell* 52:705 (1988).
- Chaiet, L., et al., *Arch. Biochem. Biophys.* 106:1 (1964).
- Chaires, J.B., et al., *Biochemistry* 29:6145-6153 (1990).
- 5 Chang, H.-K., et al., *Mol. Cell. Biol.* November:-5189-5197 (1989).
- Chen, K-X., et al., *J. Biomol. Struct. Dyn.* 3:445-466 (1985).
- 10 Chin, M.T., et al., *J. Virol.* 63:2967-2976 (1989).
- Comai, L., et al., *Cell* 68:965-976 (1992).
- Cooney, M., et al., *Science* 241:456-459 (1988).
- Courtois, G., et al., *Proc. Natl. Acad. Sci. USA* 85:7937-7941 (1988).
- 15 Cullinane, C., et al., *FEBS Lett.* 293:195-198 (1991).
- Debart, F., et al., *J. Med. Chem.* 32:1074 (1989).
- Dervan, P.B., *Science* 232:464-471 (1986).
- Descheemaeker, K.A., et al., *J. Biol. Chem.* 267(21):15086 (1992).
- 20 Edwards, C.A. et al., *J. Mol. Biol.* 180:73-90 (1984).
- Edwards, C.A., et al., in: Advances in Regulation of Cell Growth, Volume I: Regulation of Cell Growth and Activation, edited by Mond, J.J., et al., New York: Raven Press, p. 91-118 (1989).
- 25 Elias, P., et al., *Proc. Natl. Acad. Sci. USA* 85:2959-2963 (1988).
- Fox, K.R., et al., *Biochim. Biophys. Acta* 840:383-392 (1985).
- 30 Fox, K.R., et al., *Nucl. Acids Res.* 16:2489-2507 (1988).
- Fox, K.R., et al., *Nucl. Acids Res.* 18:1957-1963 (1990).
- 35 Fox, K.R., et al., *Biochem J.* 269:217-221.

- Fried, M.G., et al., *Nuc. Acid. Res.* 9:6505 (1981).
- Galas, D., et al., *Nuc. Acid Res.* 5:3157-3170 (1981).
- 5 Garner, M.M., et al., *Nuc. Acid. Res.* 9:3047 (1981).
- Gaugain, B., et al., *Biochemistry* 17:5071 (1978).
- Gessner, R.V., et al., *Biochemistry* 24:237-240 (1985).
- 10 Gilbert, D.F., et al., *Proc. Natl. Acad. Sci. USA* 86:3006 (1988)
- Gilman, A. G., et al., eds., The Pharmacological Basis of Therapeutics, Eighth Edition, Pergamon Press (1990).
- 15 Goldin, A.L., et al., *J. Virol.* 38:5-58 (1981).
- Goodisman, J., et al., *Biochemistry* 31:1046-1058 (1992).
- Green, N.M., *Adv. Protein Chem.* 29:85 (1975).
- Greenblatt, J., *Cell* 66:1067-1070 (1991).
- 20 Greene, W.C., *Annu. Rev. Immunol.* 8:453-475 (1990).
- Griffen, J.H., et al., *J. Am. Chem. Soc.* (1992).
- Griffin, L.C., et al., *Science* 245:967-971 (1989).
- Gross, D.S., et al., *Annu. Rev. Biochem.* 57:159-197 (1988).
- 25 Gurskii, G.V., et al., *Mol. Biol.* 19:177 (1985).
- Harlow, E., et al., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1988).
- Harshman, K.D., et al., *Cell* 53:321 (1988).
- 30 Hausheer, F.H., et al., *Anti-Cancer Drug Design* 5:159-167 (1990).
- Hawley, R.C., et al., *Proc. Natl. Acad. Sci. USA* 86:1105-1109 (1989).
- Helene, C., et al., *Biochim. Biophys. Acta* 35 1049:99-125 (1990).

- Helene, C., et al., *Genome* 31:413-420 (1989).
- Hoogsteen, *Acta Cryst.* 12 822 (1959).
- Innis et al., eds. PCR Protocols, a Guide to Methods and Applications, Academic Press, Inc. (1991).
- 5 Jain, S.C., et al., *J. Mol. Biol.* 68:1-20 (1972).
- Jeppesen, C., et al., *Eur. J. Biochem.* 182:437-444 (1989).
- Kadonaga, J.T., *PNAS* 83:5889-5893 (1986).
- Kissinger, K., et al., *Biochemistry* 26:5590
- 10 (1987).
- Kitadai, Y., et al., *Biochem. Biophys. Res. Commun.* 189(3):1342 (1992).
- Koff, A., et al., *J. Virol.* 62:4096-4103 (1988).
- Kotler, M., et al., *FEBS.Lett.* 21:222 (1972).
- 15 Krowicki, K., et al., *J. Org. Chem.* 52:3493 (1987).
- Kuhlmann, K.F., et al., *Nucl. Acids Res.* 5:2629 (1978).
- Laugaa, P., et al., *Biochemistry* 23:1336 (1985).
- 20 Le Pecq, J.B., et al., *Proc. Natl. Acad. Sci. U.S.A.* 72:2915-2919 (1975).
- Lee, D.K., et al., *Cell* 67:1241-1250 (1991).
- Lown, J.W., et al., *J. Org. Chem.* 50:3774 (1985).
- Lown, J.W., et al., *J. Med. Chem.* 29:1210 (1986).
- 25 Luck, G., et al., *Nucl. Acids Res.* 1:503 (1974).
- Luckow, V.A., et al., *Virology* 170:31 (1989).
- Maher III, L.J., et al., *Science* 245:725-730 (1989).
- Maher, L.J., et al., *Biochemistry* 31(1):70-81
- 30 (1992).
- Maniatis, T., et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory (1982).
- Maxam, A.M., et al., *Meth. Enzymol.*, 65:499 (1980).

- McGeoch, D.J., et al., *J. Virol.* 62:444-453 (1988).
- Meijer, I., et al., *Cell-Immunol.* 145(1):56 (1992).
- 5 Miller, et al., U.S. Patent No. 4,757,055, issued 19 July 1988.
- Montenay-Garestier, T., et al., *CIBA Found. Symp.* 158:147-157.
- Mullis, K.B., U.S. Patent No. 4,683,202, issued 28
10 July 1987.
- Mullis, K.B., et al., U.S. Patent No. 4,683,195, issued 28 July 1987.
- Nakamura, S., et al., *J. Antibiot., Ser. A.* 17:220 (1964).
- 15 Neuberger, M., et al., *Oncogene* 6(8):1325 (1991).
- Olivo, P.D., et al., *Proc. Natl. Acad. Sci. USA* 85:5414-5418 (1988).
- Olivo, P.D., et al., *J. Virology* 3:196-204 (1989).
- Pelaprat, D., et al., *J. Med. Chem.* 23:1336-1343
20 (1980).
- Perouault, L., et al., *Nature* 344:358-360 (1990).
- Phillips, D.R., *Anti Cancer Drug Design* 5:21-29 (1990).
- Phillips, et al., *Biochemistry* 29:4812-4819 (1990).
- 25 Pitha, *Biochem. Biophys. Acta* 204:39 (1970a).
- Pitha, *Biopolymers* 9:965 (1970b).
- Portugal, J., et al., *FEBS Lett.* 225:195-200 (1987).
- Quigley, G.J., et al., *Science* 232:1255-1258
30 (1986).
- Raney, A.K., et al., *J. Virol.* 66(12):6912 (1992).
- Reisman, D., et al., *Mol. Cell. Biol.* 5:1822-1832 (1985).
- Remers, W.A., Antineoplastic Agents, New York:
35 John Wiley and Sons, Inc., 1992.

- Rice, J.A., et al., *Proc. Natl. Acad. Sci. USA* 85:4158-4161 (1988).
- Ryder, K., et al., *Proc. Natl. Acad. Sci. USA* 85:1487 (1988).
- 5 Salas, X., et al., *FEBS Lett.* 292:223-228 (1991).
- Sambrook, J., et al., In Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Vol. 2 (1989).
- 10 Sanger, F., et al., *Proc Nat Acad Sci, USA*, 74:5363 (1977).
- Schmidt, A., et al., *J. Virol.* 64:4037-4041 (1990).
- Schultz, P.G., et al., *Proc. Natl. Acad. Sci. USA* 80:6834-6837 (1983).
- 15 Schuhmann, E., et al., *Allg. Microbiol.* 14:321 (1974).
- Shaw, J.P., et al., *Science* 241:202 (1988).
- Sherman, S.E., et al., *Chem. Rev.* 87:1153 (1987).
- Siebenlist, U., et al., *Proc. Natl. Acad. Sci. USA* 77:122-126 (1980).
- 20 Skorobogaty, A., et al., *Anti-Cancer Drug Design* 3:41-56 (1988).
- Smith, D.B., et al., *Gene* 67:31 (1988).
- Sobell, H.M., et al., *J. Mol. Biol.* 68:21-34 (1972).
- 25 Sobell, H.M., *Prof. Nucl. Acid. Res. Mol. Biol.* 13:153-190 (1973).
- Stow, N.D., et al., *Virology* 130:427-438 (1983).
- Stow, N.D., et al., *J. Gen. Virol.* 67:1613-1623 (1986).
- 30 Strobel, S.A., et al., *Science* 249:73-75 (1990).
- Summers, M.D., et al., A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures, Texas Agricultural Experimental Station Bulletin, No. 1555 (1987).
- 35

Summerton, J., et al., PCT International Application, Publication No. WO 86/05518, Published 25 September 1986.

Summerton, J., et al., U.S. Patent No. 5,034,506,
5 issued 23 July 1991.

Thompson, C.B., et al., *Molecular and Cell Biology*
12(3):1043 (1992).

Thrum, H., et al., Antimicrobial and Antineoplastic Chemotherapy, Prague: Czech. Med. Press, pp. 819-
10 822 (1972).

Tullius, T.D., *Ann. Rev. Biophys. Biochem.* 18:213-237 (1989).

Wang, A.H.-J., et al., *Science* 225:1115-1121 (1984).

15 Wartel, R.M., et al., *J. Biol. Chem.* 15:285-318 (1975).

Weir, H.M., et al., *Nucl. Acids Res.* 17:1409-1425 (1989).

Werner, G.H., et al., *Actual. Pharmaceut. Fr.*
20 21:133 (1963).

White, R.J., et al., *Biochemistry* 28:6259-5269 (1989).

Wirth, T., et al., *EMBO J.* 7(10):3109 (1988).

Woodbury, C.P., et al., *Biochemistry* 22(20):4730-
25 4737 (1983).

Wu, C.A., et al., *J. Virol.* 62:435-443 (1988).

Young, S.L., et al., *Proc. Natl. Acad. Sci. U.S.A.*
88:10023-10026 (1991).

Zein, N., et al., *Science* 240:1198 (1988).

30 Zimmer, C., *Pros. Nucl. Acid Res. Mol. Biol.*
15:285-318 (1975).

Background of the Invention

Several classes of small molecules that interact
35 with double-stranded DNA have been identified. Many of

these small molecules have profound biological effects. For example, many aminoacridines and polycyclic hydrocarbons bind DNA and are mutagenic, teratogenic, or carcinogenic. Other small molecules that bind DNA include: biological metabolites, some of which have applications as antibiotics and antitumor agents including actinomycin D, echinomycin, distamycin, and calicheamicin; planar dyes, such as ethidium and acridine orange; and molecules that contain heavy metals, such as cisplatin, a potent antitumor drug.

The sequence binding preferences of most known DNA binding molecules have not, to date, been identified. However, several small DNA-binding molecules have been shown to preferentially recognize specific nucleotide sequences, for example: echinomycin has been shown to preferentially bind the sequence [(A/T)CGT]/[ACG(A/T)] (Gilbert et al.); cisplatin has been shown to covalently cross-link a platinum molecule between the N7 atoms of two adjacent deoxyguanosines (Sherman et al.); and calicheamicin has been shown to preferentially bind and cleave the sequence TCCT/AGGA (Zein et al.).

Many therapeutic DNA-binding molecules (such as distamycin) that were initially identified based on their therapeutic activity in a biological screen have been later determined to bind DNA. There are several examples in the literature referring to synthetic or naturally-occurring polymers of DNA-binding drugs. Netropsin, for example, is a naturally-occurring oligopeptide that binds to the minor groove of double-stranded DNA. Netropsin contains two 4-amino-1-methylpyrrole-2-carboxylate residues and belongs to a family of similar biological metabolites from *Streptomyces* spp. This family includes distamycin, anthe-
lvencin (both of which contain three N-methylpyrrole residues), noformycin, amidomycin (both of which

contain one N-methylpyrrole residue) and kikumycin (which contains two N-methylpyrrole residues, like netropsin) (Debart, et al.). Synthetic molecules of this family have also been described, including the
5 above-mentioned molecules (Lown, et al. 1985) well as dimeric derivatives (Griffin et al., Gurskii, et al.) and certain analogues (Bialer, et al. 1980, Bialer, et al. 1981, Krowicki, et al.).

Molecules in this family, particularly netropsin
10 and distamycin, have been of interest because of their biological activity as antibacterial (Thrum et al., Schuhmann, et al.), antiparasitic (Nakamura et al.), and antiviral drugs (Becker, et al., Lown, et al. 1986, Werner, et al.).

15 Among the synthetic analogs of netropsin and distamycin are oligopeptides that have been designed to have sequence preferences different from their parent molecules. Such oligopeptides include the "lexi-
tropsin" series of analogues. The N-methylpyrrole
20 groups of the netropsin series were systematically replaced with N-methylimidazole residues, resulting in lexitropsins with increased and altered sequence specificities from the parent compounds (Kissinger, et al.). Further, a number of poly(N-methylpyrrolyl)-
25 netropsin analogues have been designed and synthesized which extend the number of residues in the oligopeptides to increase the size of the binding site (Dervan, 1986).

There are several different approaches that could
30 be taken to look for small molecules that specifically inhibit the interaction of a given DNA-binding protein with its binding sequence (cognate site). One approach would be to test biological or chemical compounds for their ability to preferentially block the binding of
35 one specific DNA:protein interaction but not others.

Such an assay would depend on the development of at least two, preferably three, DNA:protein interaction systems in order to establish controls for distinguishing between general DNA-binding molecules (polycations
5 like heparin or intercalating agents like ethidium) and DNA-binding molecules having sequence binding preferences that would affect protein/cognate binding site interactions in one system but not the other(s).

One illustration of how this system could be used
10 is as follows. Each cognate site could be placed 5' to a reporter gene (such as genes encoding β -galactoside or luciferase) such that binding of the protein to the cognate site would enhance transcription of the reporter gene. The presence of a sequence-specific
15 DNA-binding drug that blocked the DNA:protein interaction would decrease the enhancement of the reporter gene expression. Several DNA enhancers could be coupled to reporter genes, then each construct compared to one another in the presence or absence of small DNA-
20 binding test molecules. In the case where multiple protein/cognate binding sites are used for screening, a competitive inhibitor that blocks one interaction but not the others could be identified by the lack of transcription of a reporter gene in a transfected cell
25 line or in an *in vitro* assay. Only one such DNA-binding sequence, specific for the protein of interest, could be screened with each assay system. This approach has a number of limitations including limited testing capability and the need to construct the
30 appropriate reporter system for each different protein/cognate site of interest.

Another example of a system to detect sequence-specific DNA-binding molecules would involve cloning a DNA-binding protein of interest, expressing the protein
35 in an expression system (e.g., bacterial, baculovirus,

or mammalian expression systems), preparing a purified or partially purified sample of protein, then using the protein in an *in vitro* competition assay to detect molecules that blocked the DNA:protein interaction.

5 These types of systems are analogous to many receptor:ligand or enzyme:substrate screening assays developed in the past, but have the same limitations as outlined above in that a new system must be developed for every different protein/cognate site combination of
10 interest. The capacity for screening numerous different sequences is therefore limited.

Another example of a system designed to detect sequence-specific DNA-binding drugs would be the use of DNA footprinting procedures as described in the
15 literature. These methods include DNase I or other nuclease footprinting (Chaires, et al.), hydroxy radical footprinting (Portugal, et al.), methidiumpropyl EDTA(iron) complex footprinting (Schultz, et al.), photofootprinting (Jeppesen, et al.), and bidirectional
20 transcription footprinting (White, et al.). These procedures are likely to be accurate within the limits of their sequence testing capability but are seriously limited by (i) the number of different DNA sequences that can be used in one experiment (typically one test
25 sequence that represents the binding site of the DNA-binding protein under study), and (ii) the difficulty of developing high throughput screening systems.

Summary of the Invention

30 In one aspect, the invention includes a method of constructing a DNA-binding agent capable of sequence-specific binding to a duplex DNA target region. The method includes identifying in the duplex DNA, a target region containing a series of at least two non-overlapping
35 base-pair sequences of four base-pairs each, where

the four base-pair sequences are adjacent, and each sequence is characterized by sequence-preferential binding to a duplex DNA-binding small molecule. The small molecules are coupled to form a DNA-binding agent
5 capable of sequence-specific binding to said target region.

In one embodiment, the duplex-binding small molecules are identified as molecules capable of binding to a selected test sequence in a duplex DNA by
10 first adding a molecule to be screened to a test system composed of (a) a DNA-binding protein that is effective to bind to a screening sequence in a duplex DNA, with a binding affinity that is substantially independent of the test sequence adjacent the screening sequence, but
15 that is sensitive to binding of molecules to such test sequence, when the test sequence is adjacent the screening sequence, and (b) a duplex DNA having said screening and test sequences adjacent one another, where the binding protein is present in an amount that
20 saturates the screening sequence in the duplex DNA.

The test molecule is incubated in the test system for a period sufficient to permit binding of the molecule being tested to the test sequence in the duplex DNA. The degree of binding protein bound to the
25 duplex DNA before adding the test molecule is compared with that after adding the molecule. The screening sequence may be from the HSV origin of replication, and the binding protein may be UL9. Exemplary screening sequences are identified as SEQ ID NO:601, SEQ ID
30 NO:602, SEQ ID NO:615, and SEQ ID NO:641.

Specific examples of tetrameric basepair sequences include TTTC, TTTG, TTAC, TTAG, TTGC, TTGG, TTCC, TTCG, TATC, TATG, TAAC, TAAG, TAGC, TAGG, TACC, TAGC sequences. A specific example of a small molecule capable of
35 binding to these sequences is distamycin.

In another aspect, the invention includes a method of blocking transcriptional activity from a duplex DNA template. The method includes identifying in the duplex DNA, a binding site for a transcription factor and, adjacent the binding site, a target region having a series of at least two non-overlapping tetrameric base-pair sequences, where the four (tetrameric) base-pair sequences are adjacent and each sequence is characterized by sequence-preferential binding to a duplex DNA-binding small molecule. The sequences are contacted with a binding agent composed of the small molecules coupled to form a DNA-binding agent capable of sequence-specific binding to said target region.

The target may be selected, for example, from DNA sequences adjacent a binding site for a eucaryotic transcription factor, such as transcription factor TFIID, or a procaryotic transcription factor, such as transcription sigma factor.

For mammalian transcription factors, the target region is typically chosen from non-conserved regions adjacent the transcription factor binding site. Target regions can be chosen so that the small molecule binding overlaps an adjacent transcription factor DNA binding sequence (e.g., for a TFIID binding site, by 1-3 nucleotide pairs). In this case, the specificity of DNA binding for the small molecule is essentially derived from the non-conserved sequences adjacent the transcription factor binding site, in order to reduce small molecule binding at the transcription factor binding site associated with other genes.

Also disclosed is a DNA-binding agent capable of binding with base-sequence specificity to a target region in duplex DNA, where the target region contains at least two adjacent four base-pair sequences. The agent includes at least two subunits, where each

subunit is a small molecule which has a sequence-preferential binding affinity for a sequence of four base-pairs in the target region. The subunits are coupled to form a DNA-binding agent capable of sequence-specific binding to said target region.

In one general embodiment, the agent is designed for binding to a sequence in which the two tetrameric basepair sequences are separated (for example, by up to 20 basepairs, typically, 1 to 6 basepairs) and the small molecules in the agent are coupled to each other by a spacer molecule.

Also forming part of the invention is a method of constructing a binding agent capable of sequence-specific binding to a duplex DNA target region. The method includes identifying in the duplex DNA, a target region containing (i) a series of at least two adjacent non-overlapping base-pair sequences of four base-pairs each, where each four base-pair sequence is characterized by sequence-preferential binding to a duplex DNA-binding small molecule, and (ii) adjacent to (i) a DNA duplex region capable of forming a triplex with a third-strand oligonucleotide. The two small molecules are coupled to form a DNA-binding agent capable of sequence-specific binding to said target region, and the DNA-binding agent is attached to a third-strand oligonucleotide.

The binding of the DNA-binding agent to duplex DNA causes a shift from B form to A form DNA, allowing triplex binding between the third-strand polynucleotide and a portion of the target sequence.

Also disclosed is a triple-strand forming agent for use in practicing the method.

In still another aspect, the invention includes a method of ordering the sequence binding preferences a DNA-binding molecule. The method includes adding a

molecule to be screened to a test system composed of (a) a DNA-binding protein that is effective to bind to a screening sequence in a duplex DNA with a binding affinity that is substantially independent of such test sequence adjacent the screening sequence, but that is sensitive to binding of molecules to such test sequence, and (b) a duplex DNA having said screening and test sequences adjacent one another, where the binding protein is present in an amount that saturates the screening sequence in the duplex DNA. The molecule in the test system is incubated for a period sufficient to permit binding of the molecule being tested to the test sequence in the duplex DNA, and the amount of binding protein bound to the duplex DNA before and after addition of the test molecule is compared. These steps are repeated using all test sequences of interest, and the sequences are then ordered on the basis of relative amounts of protein bound in the presence of the molecule for each test sequence.

The test sequences are selected, for example, from the group of 256 possible four base sequences composed of A, G, C and T. The DNA screening sequence is preferably from the HSV origin of replication, and the binding protein is preferably UL9.

The invention also includes, a method for altering the binding characteristics of a DNA-binding protein to a duplex DNA. In the method, a binding site for the DNA-binding protein is identified in the duplex DNA and a target region identified adjacent the binding site. A small molecule is selected that is characterized by sequence-preferential binding to the target region. Such molecules can be selected by the assay and methods of the present invention. When the small molecule is bound to the target region, the small molecule is typically adjacent to the binding site for the DNA-

binding protein. Alternatively, the binding of the small molecule may overlapping the site for the DNA-binding protein by at least one nucleotide pair. In the case of such overlap, the specificity of DNA binding for the small molecule is essentially derived from non-conserved sequences adjacent the DNA-binding protein's binding site -- in order to reduce small molecule binding at similar DNA:protein binding sites at other locations. Finally,

5 the duplex DNA is contacted with the small molecule at a concentration effective to alter binding of the DNA-binding protein to its binding site.

10

In this method, contacting the duplex DNA with a small molecule can either inhibit or enhance the binding of the DNA-binding protein to its binding site: depending on the small molecule that is selected. Exemplary DNA binding proteins include DNA replication factors and a variety of transcription factors.

15

One application of this method is to eucaryotic general transcription factors (e.g., TFIID), where the target region is typically selected from DNA sequences adjacent the binding site for the eucaryotic transcription factor (e.g., SEQ ID NO:1 to SEQ ID NO:600). In one embodiment, the DNA binding protein is a eucaryotic general transcription factor and the small molecule binds, in addition to the target region, 1 to three nucleotide pairs of the DNA-binding protein's binding site. In the case of TFIID, the small molecule typically binds to (i) the target region, and (ii) up to two nucleotides of the binding site for TFIID, where the nucleotides are contiguous to the target region.

20

25

30

Generally, the present invention provides a method of screening for molecules capable of binding to a selected test sequence in a duplex DNA. In the method of the present invention a test sequence of interest is

35

selected. Such sequences can be selected, for example, from the group of sequences presented as SEQ ID NO:1 to SEQ ID NO:600. Alternatively, the test sequences can be sequences having randomly generated sequences or
5 defined sets of sequences, such as, the group of 256 possible four base sequences composed of A, G, C and T.

A duplex DNA test oligonucleotide is constructed having a screening sequence adjacent a selected test sequence, where a DNA binding protein is effective to
10 bind to the screening sequence with a binding affinity that is substantially independent of the adjacent test sequence. In such constructs the DNA protein binding to the screening sequence is sensitive to binding of test molecules to the test sequence.

15 Molecules selected for testing/screening are added to a test system composed of (a) the DNA binding protein, and (b) the duplex DNA test oligonucleotide, which contains the screening and test sequences adjacent one another. Selected molecules are incubated
20 in the test system for a period sufficient to permit binding of the molecule being tested to the test sequence in the duplex DNA. The amount of binding protein bound to the duplex DNA is compared before and after adding a test molecule. Comparison of the amount
25 of binding protein bound to the duplex DNA before and after adding a test molecule can be accomplished, for example, using a gel band-shift assay or a filter-binding assay.

In the method of the present invention a number of
30 DNA:protein interactions may be used for screening purposes. In one embodiment, the DNA screening sequence is from the HSV origin of replication and the binding protein is UL9. Exemplary HSV origin of replication screening sequences include SEQ ID NO:601,
35 SEQ ID NO:602, SEQ ID NO:615, and SEQ ID NO:641.

Other DNA:protein interactions useful in the practice of the present invention include restriction endonucleases and their cognate DNA-binding sequences. These reactions are typically carried out in the
5 absence of divalent cations.

In another embodiment, the invention includes a method of identifying test sequences in duplex DNA to which binding of a test molecule is most preferred. In this method a mixture of duplex DNA test oligonucleo-
10 tides is constructed, where each oligonucleotide has a screening sequence adjacent a test sequence as described above. The test oligonucleotides of the mixture typically contain different test sequences.

A test molecule, to be screened, is added to a
15 test reaction composed of (a) the DNA binding protein, and (b) the duplex DNA test oligonucleotide mixture. The molecule is incubated in the test reaction for a period sufficient to permit binding of the compound being tested to test sequences in the duplex DNA. Test
20 oligonucleotides are separated from test oligonucleotides bound to binding protein.

The test oligonucleotides can be separated from test oligonucleotides bound to protein by, for example, passing the test reaction through a filter, where the
25 filter is capable of capturing DNA:protein complexes but not DNA that is free of protein. One filter type useful in the practice of the present invention is the nitrocellulose filter.

The separated test oligonucleotides are then
30 amplified. These amplified test oligonucleotides are then recycled through the screening steps of the assay in order to obtain a desired degree of selection. The amplified test oligonucleotides are isolated and sequenced.

Exemplary test sequences include sequences selected from the group of 256 possible four base sequences composed of A, G, C and T. Further examples of desirable test sequences include test sequences
5 derived from the sequences presented as SEQ ID NO:1 to SEQ ID NO:600.

The amplification step in the method may be accomplished by polymerase chain reaction or other methods of amplification, including, cloning and
10 subsequent *in vivo* amplification of the cloning vector containing the sequences of interest.

These and other objects and features of the invention will be more fully appreciated when the
15 following detailed description of the invention is read in conjunction with the accompanying drawings.

Brief Description of the Figures

Figure 1A illustrates a DNA-binding protein
20 binding to a screening sequence. Figures 1B and 1C illustrate how a DNA-binding protein may be displaced or hindered in binding by a small molecule by two different mechanisms: because of steric hinderance (1B) or because of conformational (allosteric) changes
25 induced in the DNA by a small molecule (1C).

Figure 2 illustrates an assay for detecting inhibitory molecules based on their ability to preferentially hinder the binding of a DNA-binding protein to its binding site. Protein (O) is displaced from DNA
30 (//) in the presence of inhibitor (X). Two alternative capture/detection systems are illustrated, the capture and detection of unbound DNA or the capture and detection of DNA:protein complexes.

Figure 3 shows a DNA-binding protein that is able
35 to protect a biotin moiety, covalently attached to an

oligonucleotide sequence, from being recognized by streptavidin when a protein is bound to the DNA.

Figure 4 shows the incorporation of biotin and digoxigenin into a typical oligonucleotide molecule for use in the assay of the present invention. The oligonucleotide contains the binding sequence (i.e., the screening sequence) of the UL9 protein, which is underlined, and test sequences flanking the screening sequence. Figure 4 also shows the preparation of double-stranded oligonucleotides end-labeled with either digoxigenin or ³²P.

Figure 5 shows a series of sequences that have been tested in the assay of the present invention for the binding of sequence-specific small molecules.

Figure 6 outlines the clonings, into an expression vector, of a truncated form of the UL9 protein (UL9-COOH) which retains its sequence-specific DNA-binding ability.

Figure 7 shows the pVL1393 baculovirus vector containing the full length UL9 protein coding sequence.

Figure 8 is a photograph of a SDS-polyacrylamide gel showing (i) the purified UL9-COOH/glutathione-S-transferase fusion protein and (ii) the UL9-COOH polypeptide.

Figure 9 presents data demonstrating the effect on UL9-COOH binding of alterations in the test sequences that flank the UL9 screening sequence.

Figure 10A shows the effect of the addition of several concentrations of distamycin A to DNA:protein assay reactions utilizing different test sequences. Figure 10B shows the effect of the addition of actinomycin D to DNA:protein assay reactions utilizing different test sequences. Figure 10C shows the effect of the addition of Doxorubicin to DNA:protein assay reactions utilizing different test sequences.

Figure 11A illustrates a DNA capture system of the present invention utilizing biotin and streptavidin coated magnetic beads. The presence of the DNA is detected using an alkaline-phosphatase substrate that yields a chemiluminescent product. Figure 11B shows a similar reaction using biotin coated agarose beads that are conjugated to streptavidin, that in turn is conjugated to the captured DNA.

Figure 12 demonstrates a test matrix based on DNA:protein-binding data.

Figure 13 lists the top strands (5'-3') of all the possible four base pair sequences that could be used as a defined set of ordered test sequences in the assay.

Figure 14A lists the top strands (5'-3') of all the possible four base pair sequences that have the same base composition as the sequence 5'-GATC-3'. This is another example of a defined, ordered set of sequences that could be tested in the assay. Figure 14B presents the general sequence of a test oligonucleotide (SEQ ID NO:617), where XXXX is the test sequence and N = A, G, C, or T.

Figure 15 shows the results of 4 duplicate experiments in which the binding activity of distamycin was tested with all possible (256) four base pair sequences. The oligonucleotides are ranked from 1 to 256 (column 1, "rank") based on their average rank from the four experiments (column 13, "ave. rank"). (rank is shown in the first column of the chart).

Figure 16 shows the average ranks (Figure 15) plotted against the ideal ranks 1 to 256.

Figure 17 shows the average r% scores (Figure 15) plotted against the rank of 1 to 256.

Figure 18 shows the results of eight experiments with actinomycin D. The r% scores and rank are shown for each of the 256 oligonucleotides.

Figure 19 shows the average r% versus rank, by average rank (data from Figure 18).

Figure 20 shows the ideal and average ranks for each of the 256 oligonucleotides.

5 Figure 21 shows the results of a position analysis for actinomycin D preference.

Figure 22 presents the data for a dinucleotide analysis of actinomycin D binding preference.

10 Figure 23 graphically displays the results presented in Figure 22.

Figure 24 graphically displays the data presented in Figure 22, where the data are combined in a combined bar chart so that the cumulative results for any dinucleotide pair are tabulated in a single bar.

15 Figure 25 shows the top strands of 16 possible duplex DNA target sites for binding bis-distamycins.

Figure 26 shows examples of bis-distamycin target sequences for bis-distamycins with internal flexible and/or variable length linkers targeted to sites
20 comprised of two TTCC sequences, where N is any base.

Figures 27A to 27H show sample oligonucleotides for competition binding studies using the assay of the present invention.

25 Figure 28 shows the DNA sequences of the HIV proviral promoter region. Several transcription factor binding sites are marked.

Figures 29A to 29D illustrate sample test oligonucleotides for use in the polymerase chain reaction based selection technique of the present invention. In
30 Figure 29A, X is the number of bases that comprise the test site.

Figure 30 illustrates a sample test oligonucleotide for use in the assay of the present invention, where the test oligonucleotide employs several different DNA:protein interaction systems.
35

Figure 31 illustrates the results of screening a selected test sequence with a single DNA:protein interaction system. In the figure, the test site is shown in bold, the potential binding site for the test molecule is underlined.

Figure 32 illustrates the results of screening the same selected test sequence as shown in Figure 31, but using a different single DNA:protein interaction system. In the figure, the test site is shown in bold, the potential binding site for the test molecule is underlined.

Detailed Description of the Invention

I. Definitions:

Adjacent is used to describe the distance relationship between two neighboring sites. Adjacent sites are 20 or less bp apart, and can be separated by any fewer number of bases including the situation where the sites are immediately abutting one another. "Flanking" is a synonym for adjacent.

Bound DNA, as used in this disclosure, refers to the DNA that is bound by the protein used in the assay (e.g., a test oligonucleotide containing the UL9 binding sequence bound to the UL9 protein).

Coding sequences or coding regions are DNA sequences that code for RNA transcripts, unless specified otherwise.

Dissociation is the process by which two molecules cease to interact: the process occurs at a fixed average rate under specific physical conditions.

Functional binding is the noncovalent association of a protein or small molecule to the DNA molecule. In one embodiment of the assay of the present invention the functional binding of the UL9 protein to a screening sequence (i.e., its cognate DNA binding site) has

been evaluated using filter binding or gel band-shift experiments.

Half-life is herein defined as the time required for one-half of the associated complexes, e.g., DNA:protein complexes, to dissociate.

Heteropolymers are molecules comprised of at least two different subunits, each representing a different type or class of molecule. The covalent coupling of different subunits, such as, DNA-binding molecules or portions of DNA-binding molecules, results in the formation of a heteropolymer: for example, the coupling of a non-intercalating homopolymeric DNA-binding molecule, such as distamycin, to an intercalating drug, such as daunomycin. Likewise, the coupling of netropsin, which is essentially a molecular subunit of distamycin, to daunomycin would also be a heteropolymer. As a further example, the coupling of distamycin, netropsin, or daunomycin to a DNA-binding homopolymer, such as a triplex-forming oligonucleotide, would result in a heteropolymer.

Homopolymers are molecules that are comprised of a repeating subunit of the same type or class. Two examples of duplex DNA-binding homopolymers are as follows: (i) triplex-forming oligonucleotides or oligonucleotide analogs, which are composed of repeating subunits of nucleotides or nucleotide analogs, and (ii) oligopeptides, which are composed of repeating subunits linked by peptide bonds (e.g., distamycin, netropsin).

Sequence-preferential binding refers to DNA binding molecules that generally bind DNA but that show preference for binding to some DNA sequences over others. Sequence-preferential binding is typified by several of the small molecules tested in the present disclosure, e.g., distamycin. Sequence-preferential and sequence-specific binding can be evaluated using a

test matrix such as is presented in Figure 12. For a given DNA-binding molecule, there are a spectrum of differential affinities for different DNA sequences ranging from non-sequence-specific (no detectable preference) to sequence preferential to absolute sequence specificity (i.e., the recognition of only a single sequence among all possible sequences, as is the case with many restriction endonucleases).

Sequence-specific binding refers to DNA binding molecules which have a strong DNA sequence binding preference. For example, the following demonstrate typical sequence-specific DNA-binding: (i) multimers (heteropolymers and homopolymers) of the present invention (e.g., Section IV.E.1, Multimerization; Example 13), and (ii) restriction enzymes and the proteins listed in Table IV.

Screening sequence is the DNA sequence that defines the cognate binding site for the DNA binding protein: in the case of UL9, the screening sequence can, for example, be SEQ ID NO:601.

Small molecules are desirable as therapeutics for several reasons related to drug delivery, including the following: (i) they are commonly less than 10 K molecular weight; (ii) they are more likely to be permeable to cells; (iii) unlike peptides or oligonucleotides, they are less susceptible to degradation by many cellular mechanisms; and, (iv) they are not as apt to elicit an immune response. Many pharmaceutical companies have extensive libraries of chemical and/or biological mixtures, often fungal, bacterial, or algal extracts, that would be desirable to screen with the assay of the present invention. Small molecules may be either biological or synthetic organic compounds, or even inorganic compounds (i.e., cisplatin).

Test sequence is a DNA sequence adjacent the screening sequence. The assay of the present invention screens for molecules that, when bound to the test sequence, affect the interaction of the DNA-binding protein with its cognate binding site (i.e., the screening sequence). Test sequences can be placed adjacent either or both ends of the screening sequence. Typically, binding of molecules to the test sequence interferes with the binding of the DNA-binding protein to the screening sequence. However, some molecules binding to these sequences may have the reverse effect, causing an increased binding affinity of the DNA-binding protein to the screening sequence. Some molecules, even while binding in a sequence specific or sequence preferential manner, might have no effect in the assay. These molecules would not be detected in the assay.

Unbound DNA, as used in this disclosure, refers to the DNA that is not bound by the protein used in the assay (i.e., in the examples of this disclosure, the UL9 protein).

II. The Assay.

One feature of the present invention is that it provides an assay to identify small molecules that will bind in a sequence-specific manner to medically significant DNA target sites. The assay facilitates the development of a new field of pharmaceuticals that operates by interfering with specific DNA functions, such as crucial DNA:protein interactions. A sensitive, well-controlled assay has been developed (i) to detect DNA-binding molecules and (ii) to determine their sequence-specificity and affinity. The assay can be used to screen large biological and chemical libraries. For example, the assay will be used to detect sequence-

specific DNA-binding molecules in fermentation broths or extracts from various microorganisms.

Furthermore, another application for the assay is to determine the sequence specificity and relative
5 affinities of known DNA-binding drugs (and other DNA-binding molecules) for different DNA sequences. Such drugs, which are currently used primarily as antibiotics or anticancer drugs, may have previously unidentified activities that make them strong candidates for
10 therapeutics or therapeutic precursors in entirely different areas of medicine. The use of the assay to determine the sequence-binding preference of these known DNA-binding molecules enables the rational design of novel DNA-binding molecules with enhanced sequence-binding preference. The methods for designing and
15 testing these novel DNA-binding molecules is described below.

The screening assay of the present invention is basically a competition assay that is designed to test
20 the ability of a test molecule to compete with a DNA-binding protein for binding to a short, synthetic, double-stranded oligodeoxynucleotide that contains the recognition sequence for the DNA-binding protein flanked on either or both sides by a variable test
25 site. The variable test site may contain any DNA sequence that provides a reasonable recognition sequence for a DNA-binding test molecule. Molecules that bind to the test site alter the binding characteristics of the protein in a manner that can be readily detected.
30 The extent to which such molecules are able to alter the binding characteristics of the protein is likely to be directly proportional to the affinity of the test molecule for the DNA test site. The relative affinity of a given molecule for different oligonucleotide sequences at the test site (i.e., test sequences) can be
35

established by examining the molecule's effect on the DNA:protein interaction using each of the test sequences.

The assay can be used to test specific target sequences and to identify novel DNA-binding molecules. Also, the assay provides a means for the determination of the high affinity DNA binding sites for a given DNA-binding molecule, thus facilitating the identification of specific target sequences.

A. General Considerations.

The assay of the present invention has been designed for detecting test molecules or compounds that affect the rate of transfer of a specific DNA molecule from one protein molecule to another identical protein in solution.

A mixture of DNA and protein is prepared in solution. The concentration of protein is in excess to the concentration of the DNA so that virtually all of the DNA is found in DNA:protein complexes. The DNA is a double-stranded oligonucleotide that contains the recognition sequence for a specific DNA-binding protein (i.e., the screening sequence). The protein used in the assay contains a DNA-binding domain that is specific for binding to the sequence within the oligonucleotide. The physical conditions of the solution (e.g., pH, salt concentration, temperature) are adjusted such that the half-life of the complex is amenable to performing the assay (optimally a half-life of 5-120 minutes), preferably in a range that is close to normal physiological conditions.

As one DNA:protein complex dissociates, the released DNA rapidly reforms a complex with another protein in solution. Since the protein is in excess to the DNA, dissociations of one complex always result in

the rapid reassociation of the DNA into another DNA:protein complex. At equilibrium, very few DNA molecules will be unbound. If the unbound DNA is the component of the system that is measured, the minimum background of the assay is the amount of unbound DNA observed during any given measurable time period. If the capture/detection system used for capturing the unbound DNA is irreversible, the brevity of the observation period (the length of time used to capture the unbound DNA) and the sensitivity of the detection system define the lower limits of background DNA.

Figure 1 illustrates how (i) such a protein can be displaced from its cognate binding site, (ii) a protein can be prevented from binding its cognate binding site, and (iii) how the kinetics of the DNA:protein interaction can be altered. In each case, the binding site for the test molecule is located at a site flanking the recognition sequence for the DNA-binding protein (Figure 1A). One mechanism is steric hinderance of protein binding by a small molecule (competitive inhibition; Figure 1B). Alternatively, a molecule may interfere with a DNA:protein binding interaction by inducing a conformational change in the DNA (allosteric interference, noncompetitive inhibition; Figure 1C). In either event, if a test molecule that binds the oligonucleotide hinders binding of the protein, even transiently, the rate of transfer of DNA from one protein to another will be decreased. This will result in a net increase in the amount of unbound DNA and a net decrease in the amount of protein-bound DNA. In other words, an increase in the amount of unbound DNA or a decrease in the amount of bound DNA indicates the presence of an inhibitor, regardless of the mechanism of inhibition (competitive or noncompetitive).

Alternatively, molecules may be isolated that, when bound to the DNA, cause an increased affinity of the DNA-binding protein for its cognate binding site. In this case, the assay control samples (no drug added) are adjusted to less than 100% DNA:protein complex so that the increase in binding can be detected. The amount of unbound DNA (observed during a given measurable time period after the addition of the molecule) will decrease and the amount of bound DNA will increase in the reaction mixture as detected by the capture/detection system described in Section II.

B. Choosing and Testing an Appropriate DNA-Binding Protein.

Experiments performed in support of the present invention have defined an approach for identifying molecules having sequence-preferential DNA-binding. In this approach small molecules binding to sequences adjacent the cognate binding sequence can inhibit the protein/cognate DNA interaction. This assay has been designed to use a single DNA:protein interaction to screen for sequence-specific or sequence-preferential DNA-binding molecules that recognize virtually any sequence.

While DNA-binding recognition sites are usually quite small (4-17 bp), the sequence that is protected by the binding protein is larger (usually 5 bp or more on either side of the recognition sequence -- as detected by DNAase I protection (Galas, et al.) or methylation interference (Siebenlist, et al.).

Experiments performed in support of the present invention demonstrated that a single protein and its cognate DNA-binding sequence can be used to assay virtually any DNA sequence by placing a sequence of interest adjacent to the cognate site: a small mole-

5 cule bound to the adjacent site can be detected by alterations in the binding characteristics of the protein to its cognate site. Such alterations might occur by either steric hindrance (which would cause the dissociation of the protein) or induced conformational changes in the recognition sequence for the protein (which may cause either enhanced binding or, more likely, decreased binding of the protein to its cognate site).

10

1. Criteria for Choosing an Appropriate DNA-Binding Protein.

15 There are several considerations involved in choosing DNA:protein complexes that can be employed in the assay of the present invention including:

20 a.) The half-life of the DNA:protein complex should be short enough to accomplish the assay in a reasonable amount of time. The interactions of some proteins with their cognate binding sites in DNA can be measured in days not minutes: such tightly bound complexes would inconveniently lengthen the period of time it takes to perform the assay.

25 b.) The half-life of the complex should be long enough to allow the measurement of unbound DNA in a reasonable amount of time. For example, the level of free DNA is dictated by the ratio between the time needed to measure free DNA and the amount of free DNA that occurs naturally due to the dissociation of the complex during the measurement time period.

30 In view of the above two considerations, practical useful DNA:protein half-lives fall in the range of approximately two minutes to several days: shorter half-lives may be accommodated by faster equipment and longer half-lives may be accommodated by destabilizing
35 the binding conditions for the assay.

c.) A further consideration is that the kinetic interactions of the DNA:protein complex is relatively insensitive to the nucleotide sequences flanking the recognition sequence. The affinity of DNA-binding proteins may be affected by differences in the sequences adjacent to the recognition sequence. If the half-life of the complex is affected by the flanking sequence, the analysis of comparative binding data between different flanking oligonucleotide sequences becomes difficult but is not impossible.

2) Testing DNA:Protein Interactions for Use in the Assay.

a.) Other DNA:Protein Interactions Useful in the Method of the Present Invention. There are many known DNA:protein interactions that may be useful in the practice of the present invention, including (i) the DNA protein interactions listed in Table IV, (ii) bacterial, yeast, and phage systems such as lambda ϕ_L - ϕ_R /cro, and (iii) modified restriction enzyme systems (e.g., protein binding in the absence of divalent cations, see Section IV). Any protein that binds to a specific recognition sequence may be useful in the present invention. One constraining factor is the effect of the immediately adjacent sequences (the test sequences) on the affinity of the protein for its recognition sequence. DNA:protein interactions in which there is little or no effect of the test sequences on the affinity of the protein for its cognate site are preferable for use in the described assay; however, DNA:protein interactions that exhibit test-sequence-dependent differential binding may still be useful if algorithms that compensate for the differential affinity are applied to the analysis of data. In general, the effect of flanking sequence composition on

the binding of the protein is likely to be correlated to the length of the recognition sequence for the DNA-binding protein. That is, the kinetics of binding for proteins with shorter recognition sequences are more likely to suffer from flanking sequence effects, while the kinetics of binding for proteins with longer recognition sequences are more likely to not be affected by flanking sequence composition. The present disclosure provides methods and guidance for testing the usefulness of such DNA:protein interactions, in the screening assay.

b.) The Use of UL9 Proteins in the Practice of the Present Invention.

Experiments performed in support of the present invention have identified a DNA:protein interaction that is particularly useful for the above described assay: the Herpes Simplex Virus (HSV) UL9 protein that binds the HSV origin of replication (*oriS*). The UL9 protein has fairly stringent sequence specificity. There appear to be three binding sites for UL9 in *oriS*, SEQ ID NO:601, SEQ ID NO:602 and SEQ ID NO:615 (Elias, et al.; Stow, et al.). One sequence (SEQ ID NO:601) binds with at least 10-fold higher affinity than the second sequence (SEQ ID NO:602): the embodiments described below use the higher affinity binding site (SEQ ID NO:601). Another useful UL9-binding site, alibi a lower affinity binding site, SEQ ID NO:641, has also been identified.

DNA:protein association reactions are performed in solution. The DNA:protein complexes can be separated from free DNA by any of several methods. One particularly useful method for the initial study of DNA:protein interactions has been visualization of binding results using band shift gels (Example 3A). In this method DNA:protein binding reactions are applied to

polyacrylamide/TBE gels and the labelled complexes and free labeled DNA are separated electrophoretically. These gels are fixed, dried, and exposed to X-ray film. The resulting autoradiograms are examined for the amount of free probe that is migrating separately from the DNA:protein complex. These assays include (i) a lane containing only free labeled probe, and (ii) a lane where the sample is labeled probe in the presence of a large excess of binding protein. The band shift assays allow visualization of the ratios between DNA:protein complexes and free probe. However, they are less accurate than filter binding assays for rate-determining experiments due to the lag time between loading the gel and electrophoretic separation of the components.

The filter binding method is particularly useful in determining the half-life for oligonucleotide:-protein complexes (Example 3B). In the filter binding assay, DNA:protein complexes are retained on a filter while free DNA passes through the filter. This assay method is more accurate for half-life determinations because the separation of DNA:protein complexes from free probe is very rapid. The disadvantage of filter binding is that the nature of the DNA:protein complex cannot be directly visualized. So if, for example, the competing molecule was also a protein competing for the binding of a site on the DNA molecule, filter binding assays cannot differentiate between the binding of the two proteins nor yield information about whether one or both proteins are binding.

C. Preparation of Full Length UL9 and UL9-COOH Polypeptides.

UL9 protein has been prepared by a number of recombinant techniques (Example 2). The full length

UL9 protein has been prepared from baculovirus infected insect cultures (Example 3A, B, and C). Further, a portion of the UL9 protein that contains the DNA-binding domain (UL9-COOH) has been cloned into a bacterial expression vector and produced by bacterial cells (Example 3D and E). The DNA-binding domain of UL9 is contained within the C-terminal 317 amino acids of the protein (Weir, et al.). The UL9-COOH polypeptide was inserted into the expression vector in-frame with the glutathione-S-transferase (gst) protein. The gst/UL9 fusion protein was purified using affinity chromatography (Example 3E). The vector also contained a thrombin cleavage site at the junction of the two polypeptides. Therefore, once the fusion protein was isolated (Figure 8, lane 2) it was treated with thrombin, cleaving the UL9-COOH/ gst fusion protein from the gst polypeptide (Figure 8, lane 3). The UL9-COOH-gst fusion polypeptide was obtained at a protein purity of greater than 95% as determined using Coomassie staining.

Other hybrid proteins can be utilized to prepare DNA-binding proteins of interest. For example, fusing a DNA-binding protein coding sequence in-frame with a sequence encoding the thrombin site and also in-frame with the β -galactoside coding sequence. Such hybrid proteins can be isolated by affinity or immunoaffinity columns (Maniatis, et al.; Pierce, Rockford IL). Further, DNA-binding proteins can be isolated by affinity chromatography based on their ability to interact with their cognate DNA binding site. For example, the UL9 DNA-binding site (SEQ ID NO:601) can be covalently linked to a solid support (e.g., CnBr-activated Sepharose 4B beads, Pharmacia, Piscataway NJ), extracts passed over the support, the support washed, and the DNA-binding then isolated from the

support with a salt gradient (Kadonaga). Alternatively, other expression systems in bacteria, yeast, insect cells or mammalian cells can be used to express adequate levels of a DNA-binding protein for use in this assay.

The results presented below in regard to the DNA-binding ability of the truncated UL9 protein suggest that full length DNA-binding proteins are not required for the DNA:protein assay of the present invention: only a portion of the protein containing the cognate site recognition function may be required. The portion of a DNA-binding protein required for DNA-binding can be evaluated using a functional binding assay (Example 4A). The rate of dissociation can be evaluated (Example 4B) and compared to that of the full length DNA-binding protein. However, any DNA-binding peptide, truncated or full length, may be used in the assay if it meets the criteria outlined in Section II.B.1, "Criteria for choosing an appropriate DNA-binding protein". This remains true whether or not the truncated form of the DNA-binding protein has the same affinity as the full length DNA-binding protein.

D. Functional Binding and Rate of Dissociation.

The full length UL9 and purified UL9-COOH proteins were tested for functional activity in "band shift" assays (see Example 4A). The buffer conditions were optimized for DNA:protein-binding (Example 4C) using the UL9-COOH polypeptide. These DNA-binding conditions also worked well for the full-length UL9 protein. Radiolabeled oligonucleotides (SEQ ID NO:614) that contained the 11 bp UL9 DNA-binding recognition sequence (SEQ ID NO:601) were mixed with each UL9 protein in appropriate binding buffer. The reactions were incubated at room temperature for 10 minutes (binding

occurs in less than 2 minutes) and the products were separated electrophoretically on non-denaturing polyacrylamide gels (Example 4A).

5 The degree of DNA:protein-binding could be determined from the ratio of labeled probe present in DNA:protein complexes versus that present as free probe. This ratio was typically determined by optical scanning of autoradiograms and comparison of band intensities. Other standard methods may be used as
10 well for this determination, such as scintillation counting of excised bands. The UL9-COOH polypeptide and the full length UL9 polypeptide, in their respective buffer conditions, bound the target oligonucleotide equally well.

15 The rate of dissociation was determined using competition assays. An excess of unlabelled oligonucleotide that contained the UL9 binding site was added to each reaction. This unlabelled oligonucleotide acts as a specific inhibitor, capturing the UL9 protein as
20 it dissociates from the labelled oligonucleotide (Example 4B). The dissociation rate, as determined by a band-shift assay, for both full length UL9 and UL9-COOH was approximately 4 hours at 4°C or approximately 10 minutes at room temperature. Neither non-specific
25 oligonucleotides (a 10,000-fold excess) nor sheared herring sperm DNA (a 100,000-fold excess) competed for binding with the oligonucleotide containing the UL9 binding site.

30 E. oriS Flanking Sequence Variation.

As mentioned above, one feature of a DNA:protein-binding system to be used in the assay of the present invention is that the DNA:protein interaction is not affected by the nucleotide sequence of the regions
35 adjacent the DNA-binding site. The sensitivity of any

DNA:protein-binding reaction to the composition of the flanking sequences can be evaluated by the functional binding assay and dissociation assay described above.

To test the effect of flanking sequence variation
5 on UL9 binding to the *oriS* SEQ ID NO:601 sequences oligonucleotides were constructed with 20-30 different sequences (*i.e.*, the test sequences) flanking the 5' and 3' sides of the UL9 binding site. Further, oligonucleotides were constructed with point mutations at several
10 positions within the UL9 binding site. Most point mutations within the binding site destroyed recognition. Several changes did not destroy recognition and these include variations at sites that differ between the UL9 binding sites (SEQ ID NO:601, SEQ ID NO:602, SEQ ID NO:615 and SEQ ID NO:641): the second UL9
15 binding site (SEQ ID NO:602) shows a ten-fold decrease in UL9:DNA binding affinity (Elias, et al.) relative to the first (SEQ ID NO:601). On the other hand, sequence variation at the test site (also called the test sequence), adjacent to the screening site (Figure 5, Example 5), had virtually no effect on binding or the
20 rate of dissociation.

The results demonstrating that the nucleotide sequence in the test site, which flanks the screening
25 site, has no effect on the kinetics of UL9 binding in any of the oligonucleotides tested is a striking result. This allows the direct comparison of the effect of a DNA-binding molecule on test oligonucleotides that contain different test sequences. Since the
30 only difference between test oligonucleotides is the difference in nucleotide sequence at the test site(s), and since the nucleotide sequence at the test site has no effect on UL9 binding, any differential effect observed between the two test oligonucleotides in
35 response to a DNA-binding molecule must be due solely

to the differential interaction of the DNA-binding molecule with the test sequence(s). In this manner, the insensitivity of UL9 to the test sequences flanking the UL9 binding site greatly facilitates the interpretation of results. Each test oligonucleotide acts as a control sample for all other test oligonucleotides. This is particularly true when ordered sets of test sequences are tested (e.g., testing all 256 four base pair sequences (Figure 13) for binding to a single drug).

Taken together the above experiments support that the UL9-COOH polypeptide binds the SEQ ID NO:601 sequence with (i) appropriate strength, (ii) an acceptable dissociation time, and (iii) indifference to the nucleotide sequences flanking the screening site. These features suggested that the UL9/oriS system could provide a versatile assay for detection of small molecule/DNA-binding involving any number of specific nucleotide sequences.

The above-described experiment can be used to screen other DNA:protein interactions to determine their usefulness in the present assay.

F. Small Molecules as Sequence-Specific Competitive Inhibitors.

To test the utility of the present assay system several small molecules that have sequence-binding preferences (i.e., a preference for AT-rich versus GC-rich sequences) have been tested.

Distamycin A binds relatively weakly to DNA ($K_A = 2 \times 10^5 \text{ M}^{-1}$) with a preference for non-alternating AT-rich sequences (Jain, et al.; Sobell; Sobell, et al.). Actinomycin D binds DNA more strongly ($K_A = 7.6 \times 10^7 \text{ M}^{-1}$) than Distamycin A and has been reported to have a relatively strong preference for the dinucleotide se-

quence dGdC (Luck, et al.; Zimmer; Wartel). Each of these molecules poses a stringent test for the assay. Distamycin A tests the sensitivity of the assay because of its relatively weak binding. Actinomycin D challenges the ability to utilize flanking sequences since the UL9 recognition sequence contains a dGdC dinucleotide: therefore, it might be anticipated that all of the oligonucleotides, regardless of the test sequence flanking the assay site, might be equally affected by actinomycin D.

In addition, Doxorubicin, a known anti-cancer agent that binds DNA in a sequence-preferential manner (Chen, K-X, et al.), has been tested for preferential DNA sequence binding using the assay of the present invention.

Actinomycin D, Distamycin A, and Doxorubicin have been tested for their ability to preferentially inhibit the binding of UL9 to oligonucleotides containing different sequences flanking the UL9 binding site (Example 6, Figure 5). Furthermore, distamycin A and actinomycin D have been screened against all possible 256 4 bp DNA sequences. Binding assays were performed as described in Example 5. These studies were completed under conditions in which UL9 is in excess of the DNA (i.e., most of the DNA is in DNA:protein complexes).

In the preliminary studies, distamycin A was tested with 5 different test sequences flanking the UL9 screening sequence: SEQ ID NO:605 to SEQ ID NO:609. The results shown in Figure 10A demonstrate that Distamycin A preferentially disrupts binding to the test sequences UL9 polyT, UL9 polyA and, to a lesser extent, UL9 ATAT. Figure 10A also shows the concentration dependence of the inhibitory effect of distamycin A: at 1 μ M distamycin A most of the DNA:protein

complexes are intact (top band) with free probe appearing in the UL9 polyT and UL9 polyA lanes, and some free probe appearing in the UL9 ATAT lane; at 4 μ M free probe can be seen in the UL9 polyT and UL9 polyA lanes; at 16 μ M free probe can be seen in the UL9 polyT and UL9 polyA lanes; and at 40 μ M the DNA:protein in the polyT, UL9 polyA and UL9 ATAT lanes are near completely disrupted while some DNA:protein complexes in the other lanes persist. These results were consistent with the reported preference of Distamycin A for non-alternating AT-rich sequences.

Actinomycin D was tested with 8 different test sequences flanking the UL9 screening sequence: SEQ ID NO:605 to SEQ ID NO:609, and SEQ ID NO:611 to SEQ ID NO:613. The results shown in Figure 10B demonstrate that actinomycin D preferentially disrupts the binding of UL9-COOH to the oligonucleotides UL9 CCCG (SEQ ID NO:605) and UL9 GGGC (SEQ ID NO:606). These oligonucleotides contain, respectively, three or five dGdC dinucleotides in addition to the dGdC dinucleotide within the UL9 recognition sequence. This result is consistent with the results described in the literature for Actinomycin D binding to the dinucleotide sequence dGdC. Apparently the presence of a potential preferred target site within the screening sequence (*oriS*, SEQ ID NO:601), as mentioned above, does not interfere with the function of the assay.

Doxorubicin was tested with 8 different test sequences flanking the UL9 screening sequence: SEQ ID NO:605 to SEQ ID NO:609, and SEQ ID NO:611 to SEQ ID NO:613. The results shown in Figure 10C demonstrate that Doxorubicin preferentially disrupts binding to *oriEco3*, the test sequence of which differs from *oriEco2* by only one base (compare SEQ ID NO:612 and SEQ ID NO:613). Figure 10C also shows the concentration

dependence of the inhibitory effect of Doxorubicin: at 15 μ M Doxorubicin, the UL9 binding to the screening sequence is strongly affected when oriEco3 is the test sequence, and more mildly affected when polyT, UL9
5 GGGC, or oriEco2 was the test sequence; and at 35 μ M Doxorubicin most DNA:protein complexes are nearly completely disrupted, with UL9 polyT and UL9ATAT showing some DNA still complexed with protein. Also, effects similar to those observed at 15 μ M were also
10 observed using Doxorubicin at 150 nM, but at a later time point.

The feasibility studies performed with the limited set of test sequences, described above, provided evidence that the results of the assay are not inconsistent with the results reported in the literature.
15 However, the screening of all possible 256 four base-pair sequences, using the assay of the present invention, provides a much more extensive overview of the sequence preferences of distamycin A and actinomycin D.

20 The actual ranking of values obtained from the assay, for any given test compound, can be variable. A number of sequences can be clustered having similar affinity: although absolute rank might not be determinable, relative ranks can be determined.

25 The results obtained in the feasibility studies with both distamycin A and actinomycin D were corroborated by the results obtained in the screen of all 256 sequences. In other words, the rank of the oligonucleotides remained internally consistent in the larger
30 screen. Further, the screens of distamycin A and actinomycin D both support the general hypotheses described in the literature: that is, distamycin A has a preference for binding AT-rich sequences while actinomycin D has a preference for binding GC-rich
35 sequences. However, both drug screens of all possible 4

bp sequences revealed additional characteristics that have not been described in the literature.

Based on the data from 4 separate experiments (Examples 10 and 11; Figures 15, 16 and 17), consensus sequences can be derived for distamycin binding. One consensus sequence (Example 11) is relatively AT-rich, although the preference in the 4th base position is distinctly G or C. The other consensus sequence (Example 11) is relatively GC-rich, with some of the sequences having a 75% GC-content. As noted above, the assay data is consistent with distamycin binding data shown in the literature.

The ability of the assay to distinguish sequence binding preference using weak DNA-binding molecules with relatively poor sequence-specificity (such as distamycin A) is a stringent test of the assay. Accordingly, the present assay seems well-suited for the identification of molecules having better sequence specificity and/or higher sequence binding affinity. Further, the results demonstrate sequence preferential binding with the known anti-cancer drug Doxorubicin. This result indicates the assay may be useful for screening mixtures for molecules displaying similar characteristics that could be subsequently tested for anti-cancer activities as well as sequence-specific binding.

Other compounds that may be suitable for testing in the present DNA:protein system or for defining alternate DNA:protein systems include the following categories of DNA-binding molecules.

A first category of DNA-binding molecules includes non-intercalating major and minor groove DNA-binding molecules. For example, two major classes of major groove binding molecules are DNA-binding proteins (or peptides) and nucleic acids (or nucleic acid analogs

such as those with peptide or morpholino backbones) capable of forming triplex DNA. There are a number of non-intercalating minor groove DNA-binding molecules including, but not limited to the following: distamycin A, netropsin, mithramycin, chromomycin and oligomycin, which are used as antitumor agents and antibiotics; and synthetic antitumor agents such as berenil, phthalanilides, aromatic bisguanylhydrazones and bisquaternary ammonium heterocycles (for review, see Baguley, 1982). Non-intercalating DNA-binding molecules vary greatly in structure: for example, the netropsin-distamycin series are oligopeptides compared to the diarylamidines berenil and stilbamidine.

A second category of DNA-binding molecules includes intercalating DNA-binding molecules. Intercalating agents are an entirely different class of DNA-binding molecules that have been identified as antitumor therapeutics and include molecules such as daunomycin (Chaires, *et al.*) and nogalomycin (Fox, *et al.*, 1988) (see Remers, 1984).

A third category of DNA-binding molecules includes molecules that have both groove-binding and intercalating properties. DNA-binding molecules that have both intercalating and minor groove binding properties include actinomycin D (Goodisman, *et al.*), echinomycin (Fox, *et al.* 1990), triostin A (Wang, *et al.*), and luzopeptin (Fox, 1988). In general, these molecules have one or two planar polycyclic moieties and one or two cyclic oligopeptides. Luzopeptins, for instance, contain two substituted quinoline chromophores linked by a cyclic decadepsipeptide. They are closely related to the quinoxaline family, which includes echinomycin and triostin A, although they luzopeptins have ten amino acids in the cyclic peptide, while the quinoxaline family members have eight amino acids.

In addition to the major classes of DNA-binding molecules, there are also some small inorganic molecules, such as cobalt hexamine, which is known to induce Z-DNA formation in regions that contain repetitive GC sequences (Gessner, et al.). Another example is cisplatin, cis-diamminedichloroplatinum(II), which is a widely used anticancer therapeutic. Cisplatin forms a covalent intrastrand crosslink between the N7 atoms of adjacent guanosines (Rice, et al.).

Furthermore, there are a few molecules, such as calicheamicin, that have unusual biochemical structures that do not fall in any of the major categories. Calicheamicin is an antitumor antibiotic that cleaves DNA and is thought to recognize DNA sequences through carbohydrate moieties (Hawley, et al.). Several DNA-binding molecules, such as daunomycin, A447C, and cosmomycin B have sugar group, which may play a role in the recognition process.

Limited sequence preferences for some of the above drugs have been suggested: for example, echinomycin is thought to preferentially bind to the sequence (A/T)CGT (Fox, et al.). However, the absolute sequence preferences of the known DNA-binding drugs have never been demonstrated. Despite the large number of publications in this field, prior to the development of the assay described herein, no methods were available for determining sequence preferences among all possible binding sequences.

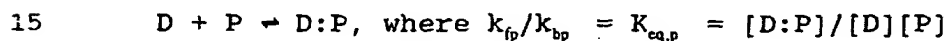
G. Theoretical Considerations on the Concentration of Assay Components.

There are two major components in the assay, the test oligonucleotide (i.e., the test sequence) and the DNA-binding domain of UL9, which is described below. A number of theoretical considerations have been

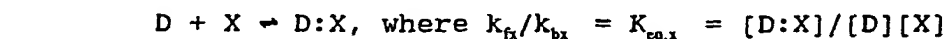
employed in establishing the assay system. In one embodiment of the invention, the assay is used as a mass-screening assay: in this embodiment the smallest volumes and concentrations possible were desirable.

5 Each assay typically uses about 0.1-0.5 ng DNA in a 15-20 μ l reaction volume (approximately 0.3-1.5 nM). The protein concentration is in excess and can be varied to increase or decrease the sensitivity of the assay. In the simplest scenario (stearic hindrance), where the

10 small molecule is acting as a competitive inhibitor and the ratio of DNA:protein and DNA-binding test molecule:DNA is 1:1, the system kinetics can be described by the following equations:



and



D = DNA, P = protein, X = DNA-binding molecule, k_{fp} and k_{fx} are the rates of the forward reaction for the DNA:protein interaction and DNA:drug interaction, respectively, and k_{bp}

25 and k_{bx} are the rates of the backwards reactions for the respective interactions. Brackets, [], indicate molar concentration of the components.

30 In the assay, both the protein, P, and the DNA-binding molecule or drug, X, are competing for the DNA. If stearic hindrance is the mechanism of inhibition, the assumption can be made that the two molecules are competing for the same site. When the concentration of

35 DNA equals the concentration of the DNA:drug or

DNA:protein complex, the equilibrium binding constant, K_{eq} , is equal to the reciprocal of the protein concentration ($1/[P]$). When all three components are mixed together, the relationship between the drug and the protein can be described as:

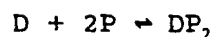
$$K_{eq,P} = z(K_{eq,X})$$

where "z" defines the difference in affinity for the DNA between P and X. For example, if $z = 4$, then the affinity of the drug is 4-fold lower than the affinity of the protein for the DNA molecule. The concentration of X, therefore, must be 4-fold greater than the concentration of P, to compete equally for the DNA molecule. Thus, the equilibrium affinity constant of UL9 will define the minimum level of detection with respect to the concentration and/or affinity of the drug. Low affinity DNA-binding molecules will be detected only at high concentrations; likewise, high affinity molecules can be detected at relatively low concentrations. With certain test sequences, complete inhibition of UL9 binding at markedly lower concentrations than indicated by these analyses have been observed, probably indicating that certain sites among those chosen for feasibility studies have affinities higher than previously published. Note that relatively high concentrations of known drugs can be utilized for testing sequence specificity. In addition, the binding constant of UL9 can be readily lowered by altering the pH or salt concentration in the assay if it ever becomes desirable to screen for molecules that are found at low concentration (e.g., in a fermentation broth or extract).

The system kinetic analysis becomes more complex if more than one protein or drug molecule is bound by

52

each DNA molecule. As an example, if UL9 binds as a dimer,



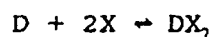
5

then the affinity constant becomes dependent on the square of the protein concentration:

$$K = [DP_2]/[D][P]^2$$

10

The same reasoning holds true for the DNA-binding test molecule, X; if,



15

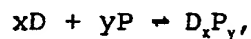
then the affinity constant becomes dependent on the square of the protein concentration:

$$K = [DX_2]/[D][X]^2$$

20

Similarly, if the molar ratio of DNA to DNA-binding test molecule was 1:3, the affinity constant would be dependent on the cube of the drug concentration.

Experimentally, the ratio of molar components can be determined. Given the chemical equation:



the affinity constant may be described as

30

$$K = [D_xP_y]/[D]^x[P]^y$$

where [] indicates concentration, D = DNA, P = protein, x = number of DNA molecules per DNA:protein complex, and y = number of protein molecules per DNA:protein

35

complex. By determining the ratio of DNA:protein complex to free DNA, one can solve for x and y:

$$\text{if } x_{\text{total}} = x_{\text{free}} + x_{\text{bound}};$$

if a = the fraction of DNA that is free, then the
5 fraction of DNA that is bound can be described as 1-a;

and if $x_{\text{bound}}:x_{\text{free}}$ (the ratio of DNA:protein complex to free DNA) is known for more than one DNA concentration. This is because the affinity constant should not vary at different DNA concentrations. Therefore,

10

$$K_{D:P, [D1]} = K_{D:P, [D2]}.$$

Substituting the right side of the equation above,

15

$$[D1_x P_y]/[D1]^x [P]^y = [D2_x P_y]/[D2]^x [P]^y.$$

Because the concentration of components in the assay can be varied and are known, the molar ratio of the components can be determined. Therefore, $[D1_x P_y]$
20 and $[D2_x P_y]$ can be described as $(1-a_1)[x_1]$ and $(1-a_2)[x_2]$, respectively, and $[D1]$ and $[D2]$ can be described as $(a_1)[x_1]$ and $(a_2)[x_2]$, respectively. $[P]$ remains constant and is described as $(y)-(y/x)(1-a)(x)$, where y is the total protein concentration and $(y/x)(1-a)(x)$ is the protein complexed with DNA.
25

The system kinetic analyses become more complex if the inhibition is allosteric (non-competitive inhibition) rather than competition by steric hindrance. Nonetheless, the probability that the relative effect
30 of an inhibitor on different test sequences is due to its relative and differential affinity to the different test sequences is fairly high. This is particularly true in the assays in which all sequences within an ordered set (e.g., possible sequences of a given length
35 or all possible variations of a certain base composi-

tion and defined length) are tested. In short, if the effect of inhibition in the assay is particularly strong for a single sequence, then it is likely that the inhibitor binds that particular sequence with higher affinity than any of the other sequences. Furthermore, while it may be difficult to determine the absolute affinity of the inhibitor, the relative affinities have a high probability of being reasonably accurate. This information will be most useful in facilitating, for instance, the refinement of molecular modeling systems.

H. The Use of the Assay under Conditions of Very High Protein Concentration.

When the screening protein is added to the assay system at very high concentrations (i.e., high enough to force binding to non-specific sites -- the protein binds to non-specific sites on the oligonucleotide as well as the screening sequence). This has been demonstrated using band shift gels: when serial dilutions are made of the protein and mixed with a fixed concentration of oligonucleotide, no binding (as seen by a band shift) is observed at very low dilutions (e.g., 1:100,000), a single band shift is observed at moderate dilutions (e.g., 1:100) and a smear, migrating higher than the single band observed at moderate dilutions, is observed at high concentrations of protein (e.g., 1:10). The observation of a smear is indicative of a mixed population of complexes, all of which presumably have the screening protein binding to the screening sequence with high affinity, but in addition have a larger number of proteins bound with markedly lower affinity to other sites.

Some of the low affinity binding proteins are likely bound to the test sequence. For example, when

using the UL9-based system, the low affinity binding proteins are likely UL9 or less likely glutathione-S-transferase: these are the only proteins in the assay mixture. These proteins are significantly more sensitive to interference by a molecule binding to the test sequence for two reasons. First, the interference is likely to be by direct steric hinderance and does not rely on induced conformational changes in the DNA; secondly, the protein is a low affinity binding protein because the test site is not a cognate-binding sequence. In the case of UL9, the difference in affinity between the low affinity binding and the high affinity binding appears to be at least two orders of magnitude.

The filter binding assays capture more DNA:protein complexes when more protein is bound to the DNA. The relative results are accurate, but under moderate protein concentrations, not all of the bound DNA (as demonstrated by band shift assays) will bind to the filter unless there is more than one DNA:protein complex per oligonucleotide (e.g., in the case of UL9, more than one UL9:DNA complex). This makes the assay exquisitely sensitive under conditions of high protein concentration. For instance, when actinomycin binds DNA at a test site under conditions where there is one DNA:UL9 complex per oligonucleotide, a preference for binding GC-rich oligonucleotides has been observed; under conditions of high protein concentration, where more than one DNA:UL9 complex is found per oligonucleotide, this binding preference is even more apparent. These results suggest that the effect of actinomycin D on a test site that is weakly bound by protein may be more readily detected than the effect of actinomycin D on the adjacent screening sequence. Therefore, employing high protein concentrations may increase the sensitivity of the assay.

III. Amplification-Based Selection Technique to Determine the Sequence Preferences of DNA-Binding Molecules.

A. Design of Test Oligonucleotides.

5 The above-described assay can be coupled to amplification methods (in one embodiment, polymerase chain reaction (Mullis, et al.; Mullis; Innis, et al.)) to achieve identification of the sequences to which binding of a test molecule is most preferred.

10 In this embodiment of the present invention, a double stranded test oligonucleotide is synthesized that contains the following elements:

(i) the binding site for a DNA-binding protein (for example, UL9), i.e., the screening site,

15 (ii) adjacent the screening site, a test site composed of more than two base pairs and preferably less than 20 base pairs (most preferably 4-12 bases), and

(iii) means to isolate selected sequences for
20 amplification, such as a sufficient number of bases flanking the test site sequences to function as priming sites for polymerase chain reaction amplification or restriction sites useful to facilitate cloning.

25 Priming sites can also be used as primer binding sites for dideoxy sequencing reactions and may contain restriction endonuclease cleavage sites to facilitate cloning manipulations.

30 The double-stranded test oligonucleotide can be generated by second-strand synthesis using a primer complementary to the priming site at the 3' end of the top-strand of the test oligonucleotide. Alternatively, both strands can be generated by other means, such as chemical synthesis, and the double-stranded test oligonucleotides can be generated by hybridization of the
35 strands.

An example of one such a test oligonucleotide is shown in Figure 29A (SEQ ID NO:630, SEQ ID NO:631 and SEQ ID NO:632). A specific example of a test oligonucleotide is shown in Figure 29B (SEQ ID NO:633), where
5 X=4. All possible 256 four base pair sequences are represented at equimolar levels within the pool of oligonucleotides generated by this sequence design.

Another example of such a test oligonucleotide sequence is shown in Figure 29C (SEQ ID NO:634), for an
10 8 base pair test sequence. In this pool of mixed sequences, all possible 8 base pair sequences ($4^8 = 65,536$) are present in equimolar amounts.

A second set of test oligonucleotides may be constructed in which the test site is placed on the
15 other side of the DNA-binding protein recognition site (e.g., Figure 29D, SEQ ID NO:635).

For any single-stranded test oligonucleotide pool, the single-stranded molecules are annealed to a primer and the bottom strands are enzymatically synthesized by
20 primer extension reactions. One advantage of using the assay/amplification PCR-cycling embodiment of the present invention is that it is convenient to work with larger test sequences in this embodiment. This protocol is geared to determining the highest affinity
25 binding sequences and is not capable of determining the rank of all test sequences nor of identifying low affinity binding sites: such ranking can be determined by screening individual sequences as described above.

30 B. Applying the Assay to the Mixed Pools of Test Oligonucleotides.

Using double-stranded test oligonucleotides, such as those just described, the basic assay is performed essentially as described above (Section I): typically
35 without the use of radioactive detection systems. As

previously discussed, a number of DNA:protein interactions may be used in this assay system. One example of such a system is the interaction of the DNA-binding domain of UL9 (or UL9-COOH) with its cognate recognition sequence.

In this embodiment of the present invention, UL9-COOH is added to the test oligonucleotide pool (for example, 256 four base pair sequences are represented at equimolar levels within the pool of oligonucleotides described above) in UL9 binding buffer. DNA-binding molecules are tested for the ability to differentially disrupt the binding of the UL9 DNA:protein complex by binding to the test sequence. After the addition of the test molecule or test mixture (e.g., a fermentation broth or fungal extract), the assay mixture is incubated for a desired time, then passed through a nitrocellulose filter. DNA:protein (such as DNA:UL9) complexes are captured on the filter. DNA that is not bound by protein passes through the filter (i.e., the filtrate) (step 1). The volume of the assay is adjusted to accommodate the amount required for the filtering process: that is, taking into consideration the losses incurred during the filtering process.

25 C. Amplification.

In one embodiment, the DNA present in the filtrate is amplified using the polymerase chain reaction (PCR) technology (Mullis; Mullis, et al.; Perkin Elmer-Cetus). An aliquot of the resulting PCR-amplified material is cycled through the DNA:protein binding assay again (step 2), then PCR-amplified again (step 3). Steps 1-3 are repeated several times using each subsequent filtrate. After each PCR amplification, part of the PCR-amplified material is retained for sequencing analysis. The result of the repeated

cyclings through the assay/amplification process is that the test oligonucleotide sequences that are amplified contain test sequences that are preferred binding sites for the test molecules. Through subsequent rounds of assay/amplification, these oligonucleotides are amplified to represent a larger and larger percent of the total population of amplified DNA molecules.

In addition to PCR, the DNA present in the filtrate can be amplified by other methods as well. For example, the DNA present in the filtrate can be cloned into a selected vector (such as, phage vectors, e.g., lambda-based, or standard cloning vectors, e.g., pBR322- or pUC-based). The cloned sequences are then transformed into an appropriate host organism in which the selected vector can replicate (for example, bacteria or yeast). The transformed host organism is cultured with concurrent amplification of the vectors containing the cloned sequences. The vectors are then isolated by standard procedures (Maniatis, et al.; Sambrook, et al.; Ausubel, et al.). Typically, the cloned sequences, originally obtained from the DNA filtrate, are obtained from the vector by restriction endonuclease digestion and size-fractionation (for example, electrophoretic separation of the digestion products followed by electroelution of the cloned sequences of interest) (Ausubel, et al.). These isolated amplified test oligonucleotide sequences can then be recycled through subsequent rounds of assay/amplification as described above.

In another embodiment, the oligonucleotide sequences present in the original DNA filtrate can be isolated, sequenced and amplified by *in vitro* synthesis of copies of the oligonucleotides.

D. Sequencing of Amplified DNA.

Samples from each cycle are sequenced using, for example, radio-labeled primers and dideoxy sequencing methodologies (Sanger) or the chemical methodologies outlined by Maxam and Gilbert. If the amplified sequences are not sufficiently resolved to obtain a unambiguous sequence information, then the DNA is further purified and sequenced. For example, the DNA is cleaved at the restriction endonuclease sites within the primer sequences and subcloned into a convenient sequencing vector, such as "BLUESCRIPT" (Stratagene, La Jolla, CA). The sequencing vectors carrying the amplified inserts are transformed into bacteria. The resulting cloned vectors are isolated and sequenced (in the case of "BLUESCRIPT," using the commercially available primers and protocols).

IV. Modifications of Test Oligonucleotides and other Useful DNA:Protein Interactions

One class of DNA:protein interactions that may be useful in the assay of the present invention is the restriction endonuclease:restriction site class of DNA:protein interactions. In the absence of divalent cations, restriction endonucleases bind DNA but have no enzymatic activity (cleavage of DNA does not take place without divalent cations). This allows the assay of the present invention to be performed using a restriction endonuclease with its cognate binding site as the screening sequence. The use of the restriction endonuclease:restriction site interaction as the basis of the present assay is described in greater detail in Section VI.B.4(c).

The test oligonucleotides of the present invention can be modified to contain two different DNA:protein screening systems, i.e., two different screening se-

quences with their respective cognate binding proteins. In the assay described above, the UL9 screening sequence lies on one side of and immediately adjacent to the test sequence. A second screening sequence, such as, a restriction endonuclease recognition sequence (restriction site), can be introduced immediately adjacent to the other side of the test sequence.

Several restriction enzymes may recognize the same restriction site. These enzymes are not identical, however, in that the cleavage sites may be at the 5' end, the center, or the 3' end of the recognition sequence. For this reason, a restriction site that is recognized by more than one restriction enzyme may be incorporated adjacent to the test site. This allows a single pool of test oligonucleotides to be used in assays employing three different DNA:protein interactions: the screening sequence has the same sequence but the restriction endonuclease used in the assay system is different in each case. Using this method to design test oligonucleotides, the UL9 screening sequence may be placed on one side of a test sequence and a restriction site screening sequence (having three cognate binding proteins) may be placed on the other side of the test sequence. Such a test oligonucleotide construction allows 4 different DNA:protein assay interaction systems to be employed with a single pool of test sequences.

One example of test oligonucleotides using several different DNA:protein interaction systems are shown in Figure 30. The top strands of the pool of test oligonucleotides shown in Figure 30 have 6 base pair test sequences (NNNNNN) and represent synthetic pools of all possible 4096 test sequences. The remainder of the nucleotide sequence is fixed. The test oligonucleotides contain the UL9 recognition sequence, 5'-

CGTTCGCACTT-3' (underlined) on one side of the test sequence and a restriction endonuclease binding sequence, 5'-GGTACC-3' (bold), on the other side of the test site. The restriction endonuclease recognition sequence is recognized by the three different restriction endonucleases *Asp718*, *RsaI* and *KpnI*. In Figure 30 the UL9 binding site (screening sequence) is located 3' of the test sequence: the UL9 binding site (screening sequence) can also be located 5' of the test sequence.

10 The shorter sequences shown above the 5' and 3' ends of the test oligonucleotides are primer sequences useful for sequencing and PCR amplification. The primer sequences contain commonly used restriction endonuclease sites for the purpose of subcloning into sequencing vectors.

15

Performing the assay with two or more different protein/screening sequence systems allows the confirmation of putative high affinity binding between a test compound and specific test sequences.

20 Alternatively, since there is no assurance that a test molecule that binds the test sequence will have the same effect on protein binding at both adjacent flanking sequences, simultaneous use of both test systems may reduce the number of false negatives detected in an assay. For example, a test molecule that does not affect the binding of protein at one screening site but may effect the binding of a different protein at the other screening site.

25

30 V. Capture/Detection Systems.

As an alternative to the above described band shift gels and filter binding assays, the measurement of inhibitors can be monitored by measuring either the level of unbound DNA in the presence of test molecules or mixtures or the level of DNA:protein complex

35

remaining in the presence of test molecules or mixtures. Measurements may be made either at equilibrium or, in a kinetic assay, prior to the time at which equilibrium is reached. The type of measurement is likely to be dictated by practical factors, such as the length of time to equilibrium, which will be determined by both the kinetics of the DNA:protein interaction as well as the kinetics of the DNA:drug interaction. The results (*i.e.*, the detection of DNA-binding molecules and/or the determination of their sequence preferences) should not vary with the type of measurement taken (kinetic or equilibrium).

Figure 2 illustrates an assay for detecting inhibitory molecules based on their ability to preferentially hinder the binding of a DNA-binding protein. In the presence of an inhibitory molecule (X) the equilibrium between the DNA-binding protein and its binding site (screening sequence) is disrupted. The DNA-binding protein (O) is displaced from DNA (/) in the presence of inhibitor (X), the DNA free of protein or, alternatively, the DNA:protein complexes, can then be captured and detected.

For maximum sensitivity, unbound DNA and DNA:protein complexes should be sequestered from each other in an efficient and rapid manner. The method of DNA capture should allow for the rapid removal of the unbound DNA from the protein-rich mixture containing the DNA:protein complexes.

Even if the test molecules are specific in their interaction with DNA they may have relatively low affinity and they may also be weak binders of non-specific DNA or have non-specific interactions with DNA at low concentrations. In either case, their binding to DNA may only be transient, much like the transient binding of the protein in solution. Accordingly, one

feature of the assay is to take a molecular snapshot of the equilibrium state of a solution comprised of the test oligonucleotide DNA, the protein, and the inhibitory test molecule. In the presence of an inhibitor, the amount of DNA that is not bound to protein will be greater than in the absence of an inhibitor. Likewise, in the presence of an inhibitor, the amount of DNA that is bound to protein will be lesser than in the absence of an inhibitor.

Any method used to separate the DNA:protein complexes from unbound DNA, should be rapid, because when the capture system is applied to the solution (if the capture system is irreversible), the ratio of unbound DNA to DNA:protein complex will change at a predetermined rate, based purely on the off-rate of the DNA:protein complex. This step, therefore, determines the limits of background. Unlike the protein and inhibitor, the capture system should bind rapidly and tightly to the DNA or DNA:protein complex. The longer the capture system is left in contact with the entire mixture of unbound DNA and DNA:protein complexes in solution, the higher the background, regardless of the presence or absence of inhibitor.

Two exemplary capture systems are described below for use in the assay of the present invention. One capture system has been devised to capture unbound DNA (Section V.A). The other has been devised to capture DNA:protein complexes (Section V.B). Both systems are amenable to high throughput screening assays. The same detection methods (Section V.C) can be applied to molecules captured using either capture system.

A. Capture of Unbound DNA.

One capture system that has been developed in the course of experiments performed in support of the

present invention utilizes a streptavidin/biotin interaction for the rapid capture of unbound DNA from the protein-rich mixture, which includes unbound DNA, DNA:protein complexes, excess protein and the test molecules or test mixtures. Streptavidin binds with extremely high affinity to biotin ($K_d = 10^{-15}M$) (Chaiet, et al.; Green). Accordingly, two advantages of the streptavidin/biotin system are that binding between the two molecules can be rapid and the interaction is the strongest known non-covalent interaction.

In this detection system a biotin molecule is covalently attached in the oligonucleotide screening sequence (i.e., the DNA-binding protein's binding site). This attachment is accomplished in such a manner that the binding of the DNA-binding protein to the DNA is not destroyed. Further, when the protein is bound to the biotinylated sequence, the protein prevents the binding of streptavidin to the biotin. In other words, the DNA-binding protein is able to protect the biotin from being recognized by the streptavidin. This DNA:protein interaction is illustrated in Figure 3.

The capture system is described herein for use with the UL9/oriS system described above. The following general testing principles can, however, be applied to analysis of other DNA:protein interactions. The usefulness of this system depends on the biophysical characteristics of the particular DNA:protein interaction.

1. Modification of the Protein Recognition Sequence with Biotin.

The recognition sequence for the binding of the UL9 (Koff, et al.) protein is underlined in Figure 4. Oligonucleotides were synthesized that contain the

UL9 binding site and site-specifically biotinylated a number of locations throughout the binding sequence (SEQ ID NO:614; Example 1, Figure 4). These biotinylated oligonucleotides were then used in band shift assays to determine the ability of the UL9 protein to bind to the oligonucleotide. These experiments using the biotinylated probe and a non-biotinylated probe as a control demonstrate that the presence of a biotin at the #8-T (biotinylated deoxyuridine) position of the bottom strand meets the requirements listed above: the presence of a biotin moiety at the #8 position of the bottom strand does not markedly affect the specificity of UL9 for the recognition site. Further, in the presence of bound UL9, streptavidin does not recognize the presence of the biotin moiety in the oligonucleotide. Biotinylation at other A or T positions did not have the two necessary characteristics (i.e., UL9 binding and protection from streptavidin): biotinylation at the adenosine in position #8, of the top strand, prevented the binding of UL9; biotinylation of either adenosines or thymidines (top or bottom strand) at positions #3, #4, #10, or #11 all allowed binding of UL9, but in each case, streptavidin also was able to recognize the presence of the biotin moiety and thereby bind the oligonucleotide in the presence of UL9.

The above result (the ability of UL9 to bind to an oligonucleotide containing a biotin within the recognition sequence and to protect the biotin from streptavidin) was unexpected in that methylation interference data (Koff, et al.) suggest that methylation of the deoxyguanosine residues at positions #7 and #9 of the recognition sequence (on either side of the biotinylated deoxyuridine) blocks UL9 binding. In these methylation interference experiments, guanosines are methylated by dimethyl sulfate at the N⁷ position,

which corresponds structurally to the 5-position of the pyrimidine ring at which the deoxyuridine is biotinylated. These moieties all protrude into the major groove of the DNA. The methylation interference data suggest
5 that the #7 and #9 position deoxyguanosines are contact points for UL9, it was therefore unexpected that the presence of a biotin moiety between them would not interfere with binding.

The binding of the full length protein was
10 relatively unaffected by the presence of a biotin at position #8 within the UL9 binding site. The rate of dissociation was similar for full length UL9 with both biotinylated and un-biotinylated oligonucleotides. However, the rate of dissociation of the truncated UL9-
15 COOH polypeptide was faster with the biotinylated oligonucleotides than with non-biotinylated oligonucleotides (for non-biotinylated oligonucleotides the rate comparable to that of the full length protein with either DNA).

20 The binding conditions were optimized for UL9-COOH so that the half-life of the truncated UL9 from the biotinylated oligonucleotide was 5-10 minutes (optimized conditions are given in Example 4), a rate compatible with a mass screening assay. The use of
25 multi-well plates to conduct the DNA:protein assay of the present invention is one approach to mass screening.

30 2. Capture of Site-Specific Biotinylated Oligonucleotides.

The streptavidin:biotin interaction can be employed in several different ways to remove unbound DNA from the solution containing the DNA, protein, and test molecule or mixture. Magnetic polystyrene or
35 agarose beads, to which streptavidin is covalently

attached or attached through a covalently attached biotin, can be exposed to the solution for a brief period, then removed by use, respectively, of magnets or a filter mesh. Magnetic streptavidinated beads are currently the method of choice. Streptavidin has been used in many of these experiments, but avidin is equally useful.

An example of a second method for the removal of unbound DNA is to attach streptavidin to a filter by first linking biotin to the filter, binding streptavidin, then blocking nonspecific protein binding sites on the filter with a nonspecific protein such as albumin. The mixture is then passed through the filter, unbound DNA is captured and the bound DNA passes through the filter. This method can give high background due to partial retention of the DNA:protein complex on the filter.

One convenient method to sequester captured DNA is the use of streptavidin-conjugated superparamagnetic polystyrene beads as described in Example 7. These beads are added to the assay mixture to capture the unbound DNA. After capture of DNA, the beads can be retrieved by placing the reaction tubes in a magnetic rack, which sequesters the beads on the reaction chamber wall while the assay mixture is removed and the beads are washed. The captured DNA is then detected using one of several DNA detection systems, as described below.

Alternatively, avidin-coated agarose beads can be used. Biotinylated agarose beads (immobilized D-biotin, Pierce) are bound to avidin. Avidin, like streptavidin, has four binding sites for biotin. One of these binding sites is used to bind the avidin to the biotin that is coupled to the agarose beads via a 16 atom spacer arm: the other biotin binding sites

remain available. The beads are mixed with binding mixtures to capture biotinylated DNA (Example 7). Alternative methods (Harlow, et al.) to the bead capture methods just described include the following
5 streptavidinated or avidinated supports: low-protein binding filters, or 96-well plates.

B. Capture of DNA:Protein Complexes.

The amount of DNA:protein complex remaining in the
10 assay mixture in the presence of an inhibitory molecule can also be determined as a measure of the relative effect of the inhibitory molecule. A net decrease in the amount of DNA:protein complex in response to a test molecule is an indication of the presence of an
15 inhibitor. DNA molecules that are bound to protein can be captured on nitrocellulose filters. Under low salt conditions, DNA that is not bound to protein freely passes through the filter. Thus, by passing the assay mixture rapidly through a nitrocellulose filter, the
20 DNA:protein complexes and unbound DNA molecules can be rapidly separated. This has been accomplished on nitrocellulose discs using a vacuum filter apparatus or on slot blot or dot blot apparatuses (all of which are available from Schleicher and Schuell, Keene, NH). The
25 assay mixture is applied to and rapidly passes through the wetted nitrocellulose under vacuum conditions. Any apparatus employing nitrocellulose filters or other filters capable of retaining protein while allowing free DNA to pass through the filter would be suitable
30 for this system.

C. Detection Systems.

For either of the above capture methods, the amount of DNA that has been captured is quantitated.
35 The method of quantitation depends on how the DNA has

been prepared. If the DNA is radioactively labelled, beads can be counted in a scintillation counter, or autoradiographs can be taken of dried gels or nitrocellulose filters. The amount of DNA has been quantitated
5 in the latter case by a densitometer (Molecular Dynamics, Sunnyvale, CA); alternatively, filters or gels containing radiolabeled samples can be quantitated using a phosphorimager (Molecular Dynamics). Further, the captured DNA may be detected using a chemilum-
10 inescent or colorimetric detection system.

Radiolabelling and chemiluminescence (i) are very sensitive, allowing the detection of sub-femtomole quantities of oligonucleotide, and (ii) use well-established techniques. In the case of chemiluminescent detection, protocols have been devised to
15 accommodate the requirements of a mass-screening assay. Non-isotopic DNA detection techniques have principally incorporated alkaline phosphatase as the detectable label given the ability of the enzyme to give a high
20 turnover of substrate to product and the availability of substrates that yield chemiluminescent or colored products.

1. Radioactive Labeling.

25 Many of the experiments described above for UL9 DNA:protein-binding studies have made use of radio-labelled oligonucleotides. The techniques involved in radiolabelling of oligonucleotides have been discussed above. A specific activity of 10^8 - 10^9 dpm per μ g DNA
30 is routinely achieved using standard methods (e.g., end-labeling the oligonucleotide with adenosine γ - $[^{32}\text{P}]$ -5' triphosphate and T4 polynucleotide kinase). This level of specific activity allows small amounts of DNA to be measured either by autoradiography of gels or

filters exposed to film or by direct counting of samples in scintillation fluid.

2. Chemiluminescent Detection.

5 For chemiluminescent detection, digoxigenin-labelled oligonucleotides (Example 1) can be detected using the chemiluminescent detection system "SOUTHERN LIGHTS," developed by Tropix, Inc. (Bedford, MA). The detection system is diagrammed in Figures 11A and 11B.

10 The technique can be applied to detect DNA that has been captured on either beads, filters, or in solution.

Alkaline phosphatase is coupled to the captured DNA without interfering with the capture system. To do this several methods, derived from commonly used ELISA

15 (Harlow, et al.; Pierce, Rockford IL) techniques, can be employed. For example, an antigenic moiety is incorporated into the DNA at sites that will not interfere with (i) the DNA:protein interaction, (ii) the DNA:drug interaction, or (iii) the capture system.

20 In the UL9 DNA:protein/biotin system the DNA has been end-labelled with digoxigenin-11-dUTP (dig-dUTP) and terminal transferase (Example 1, Figure 4). After the DNA was captured and removed from the DNA:protein mixture, an anti-digoxigenin-alkaline phosphatase

25 conjugated antibody was then reacted (Boehringer Mannheim, Indianapolis IN) with the digoxigenin-containing oligonucleotide. The antigenic digoxigenin moiety was recognized by the antibody-enzyme conjugate. The presence of dig-dUTP altered neither the ability of

30 UL9-COOH protein to bind the *oriS* (SEQ ID NO:601)-containing DNA nor the ability of streptavidin to bind the incorporated biotin.

Captured DNA was detected using the alkaline phosphatase-conjugated antibodies to digoxigenin as

35 follows. One chemiluminescent substrate for alkaline

phosphatase is 3-(2'-spiroadamantane)-4-methoxy-4-(3"-phosphoryloxy) phenyl-1,2-dioxetane disodium salt (AMPPD) (Example 7). Dephosphorylation of AMPPD results in an unstable compound, which decomposes, releasing a prolonged, steady emission of light at 477 nm. Light measurement is very sensitive and can detect minute quantities of DNA (e.g., 10^2 - 10^3 attomoles) (Example 7).

Colorimetric substrates for the alkaline phosphatase system have also been tested. While the colorimetric substrates are useable in the present assay system, use of the light emission system is more sensitive.

An alternative to the above biotin capture system is to use digoxigenin in place of biotin to modify the oligonucleotide at a site protected by the DNA-binding protein at the assay site: biotin is then used to replace the digoxigenin moieties in the above described detection system. In this arrangement the anti-digoxigenin antibody is used to capture the oligonucleotide probe when it is free of bound protein. Streptavidin conjugated to alkaline phosphatase is then used to detect the presence of captured oligonucleotides.

25 D. Alternative Methods for Detecting Molecules that Increase the Affinity of the DNA-Binding Protein for its Cognate Site.

In addition to identifying molecules or compounds that cause a decreased affinity of the DNA-binding protein for the screening sequence, molecules may be identified that increase the affinity of the protein for its cognate binding site. In this case, leaving the capture system for unbound DNA in contact with the assay for increasing amounts of time allows the establishment of a fixed half-life for the DNA:protein

complex (for example, using SEQ ID NO:601/UL9). In the presence of a stabilizing molecule, the half-life, as detected by the capture system time points, will be shortened.

- 5 Using the capture system for DNA:protein complexes to detect molecules that increase the affinity of the DNA-binding protein for the screening sequence requires that an excess of unlabeled oligonucleotide containing the UL9 binding site (but not the test sequences) is
10 added to the assay mixture. This is, in effect, an off-rate experiment. In this case, the control sample (no test molecules or mixtures added) will show a fixed off-rate. For example, samples would be taken at fixed intervals after the addition of the unlabeled competi-
15 tion DNA molecule, applied to nitrocellulose, and a decreasing amount of radiolabeled DNA:protein complex would be observed). In the presence of a DNA-binding test molecule that enhanced the binding of UL9, the off-rate would be decreased (i.e., the amount of
20 radiolabeled DNA:protein complexes observed would not decrease as rapidly at the fixed time points as in the control sample).

VI. Utility.

- 25 A. The Usefulness of Sequence-Specific DNA-Binding Molecules.

 The present invention defines a high through-put *in vitro* screening assay to test large libraries of biological or chemical mixtures for the presence of
30 DNA-binding molecules having sequence binding preference. The assay is also capable of determining the sequence-specificity and relative affinity of known DNA-binding molecules or purified unknown DNA-binding molecules. Sequence-specific DNA-binding molecules are of
35 particular interest for several reasons, which are

listed here. These reasons, in part, outline the rationale for determining the usefulness of DNA-binding molecules as therapeutic agents:

First, for a given DNA:protein interaction, there
5 are generally several thousands fewer target DNA-binding sequences per cell than protein molecules that bind to the DNA. Accordingly, even fairly toxic molecules might be delivered in sufficiently low concentration to exert a biological effect by binding to the
10 target DNA sequences.

Second, DNA has a relatively more well-defined structure compared to RNA or protein. Since the general structure of DNA has less tertiary structural variation, identifying or designing specific binding
15 molecules should be easier for DNA than for either RNA or protein. Double-stranded DNA is a repeating structure of deoxyribonucleotides that stack atop one another to form a linear helical structure. In this manner, DNA has a regularly repeating "lattice"
20 structure that makes it particularly amenable to molecular modeling refinements and hence, drug design and development.

Third, since many single genes (*i.e.*, genes which have only 1 or 2 copies in the cell) are transcribed
25 into more than one, potentially as many as thousands of RNA molecules, each of which may be translated into many proteins, targeting any DNA site, whether it is a regulatory sequence, non-coding sequence or a coding sequence, may require a much lower drug dose than
30 targeting RNAs or proteins. Proteins (*e.g.*, enzymes, receptors, or structural proteins) are currently the targets of most therapeutic agents. More recently, RNA molecules have become the targets for antisense or ribozyme therapeutic molecules.

Fourth, blocking the function of a RNA that encodes a protein or of the protein itself when that protein regulates several cellular genes may have detrimental effects: particularly if some of the regulated genes are important for the survival of the cell. However, blocking a DNA-binding site that is specific to a single gene regulated by such a protein results in reduced toxicity.

An example situation is HNF-1 binding to Hepatitis B virus (HBV): HNF-1 binds an HBV enhancer sequence and stimulates transcription of HBV genes (Chang, et al.). In a normal cell HNF-1 is a nuclear protein that appears to be important for the regulation of many genes, particularly liver-specific genes (Courtois, et al.). If molecules were isolated that specifically bound to the DNA-binding domain of HNF-1, all of the genes regulated by HNF-1 would be down-regulated, including both viral and cellular genes. Such a drug could be lethal since many of the genes regulated by HNF-1 may be necessary for liver function. However, the assay of the present invention presents the ability to screen for a molecule that could distinguish the HNF-1 binding region of the Hepatitis B virus DNA from cellular HNF-1 sites by, for example, including divergent flanking sequences when screening for the molecule. Such a molecule would specifically block HBV expression without effecting cellular gene expression.

B. General Applications of the Assay.

General applications of the assay include but are not limited to: screening libraries of unknown chemicals, either biological or synthetic compounds, for sequence-specific DNA-binding molecules, determining the sequence-specificity or preference and/or relative affinities of DNA-binding molecules, testing

of modified derivatives of DNA-binding molecules for altered specificity or affinity, using the assay in secondary confirmatory or mechanistic experiments, using the data generated from the above applications to
5 refine the predictive capabilities of molecular modeling systems, and using the refined molecular modeling systems to generate a new "alphabet" of DNA-binding subunits that can be polymerized to make novel heteropolymers designed de novo to bind specific DNA
10 target sites.

1. Mass-Screening of Libraries for the Presence of Sequence-Specific DNA-Binding Molecules.

15 Many organizations (e.g., the National Institutes of Health, pharmaceutical and chemical corporations) have large libraries of chemical or biological compounds from synthetic processes or fermentation broths or extracts that may contain as yet
20 unidentified DNA-binding molecules. One utility of the assay is to apply the assay system to the mass-screening of these libraries of different broths, extracts, or mixtures to detect the specific samples that contain the DNA-binding molecules. Once the specific mixtures
25 that contain the DNA-binding molecules have been identified, the assay has a further usefulness in aiding in the purification of the DNA-binding molecule from the crude mixture. As purification schemes are applied to the mixture, the assay can be used to test
30 the fractions for DNA-binding activity. The assay is amenable to high throughput (e.g., a 96-well plate format automated on robotics equipment such as a Beckman Biomek workstation [Beckman, Palo Alto, CA] with detection using semi-automated plate-reading
35 densitometers, luminometers, or phosphoimagers).

The concentration of protein used in mass-screening is determined by the sensitivity desired. The screening of known compounds, as described in Section VI.B.2, is typically performed in protein excess at a protein concentration high enough to produce 90-95% of the DNA bound in DNA:protein complex. The assay is very sensitive to discriminatory inhibition at this protein concentration. For some mass-screening, it may be desirable to operate the assay under higher protein concentration, thus decreasing the sensitivity of the assay so that only fairly high affinity molecules will be detected: for example, when screening fermentation broths with the intent of identifying high affinity binding molecules. The range of sensitivities in the assay will be determined by the absolute concentration of protein used.

One utility of the method of the present invention, under conditions using a relatively insensitive system (high [P]:[D] ratio), is as a screening system for novel restriction enzymes. In this case, an ability to discriminate between slight differences in affinity to different sequences may not be necessary or desirable. Restriction enzymes have highly discriminatory recognition properties -- the affinity constant of a restriction endonuclease for its specific recognition sequence versus non-specific sequences are orders of magnitude different from one another. The assay may be used to screen bacterial extracts for the presence of novel restriction endonucleases. The 256 test oligonucleotides described in Example 10, for example, may be used to screen for novel restriction endonucleases with 4 bp recognition sequences. The advantages of the system are that all possible 4 bp sequences are screened simultaneously, that is, it is not limited to self-complementary sequences. Further, any lack of

specificity (such as, more than one binding site) is uncovered during the primary screening assay.

2. Directed Screening.

5 The assay of the present invention is also useful for screening molecules that are currently described in the literature as DNA-binding molecules but with uncertain DNA-binding sequence specificity (i.e., having either no well-defined preference for
10 binding to specific DNA sequences or having certain higher affinity binding sites but without defining the relative preference for all possible DNA binding sequences). The assay can be used to determine the specific binding sites for DNA-binding molecules, among
15 all possible choices of sequence that bind with high, low, or moderate affinity to the DNA-binding molecule. Actinomycin D, Distamycin A, and Doxorubicin (Example 6) all provide examples of molecules with these modes of binding. Many anti-cancer drugs, such as Doxorubi-
20 cin (see Example 6), show binding preference for certain identified DNA sequences, although the absolute highest and lowest specificity sequences have yet to be determined, because, until the invention described herein, methods (Salas and Portugal; Cullinane and
25 Phillips; Phillips; and Phillips, et al.) for detecting differential affinity DNA-binding sites for any drug were limited. Doxorubicin is one of the most widely used anti-cancer drugs currently available. As shown in Example 6, Doxorubicin is known to bind some se-
30 quences preferentially. Another example of such sequence binding preference is Daunorubicin (Chen, et al.) which differs slightly in structure from Doxorubicin (Goodman, et al.). Both Daunorubicin and Doxorubicin are members of the anthracycline antibiotic
35 family: antibiotics in this family, and their deriva-

tives, are among the most important newer antitumor agents (Goodman, et al.).

The assay of the present invention allows the sequence preferences or specificities of DNA-binding molecules to be determined. The DNA-binding molecules for which sequence preference or specificity can be determined may include small molecules such as amino-acridines and polycyclic hydrocarbons, planar dyes, various DNA-binding antibiotics and anticancer drugs, as well as DNA-binding macromolecules, such as, peptides and polymers that bind to nucleic acids (e.g., DNA and the derivatized homologs of DNA that bind to the DNA helix).

The molecules that can be tested in the assay for sequence preference/specificity and relative affinity to different DNA sites include both major and minor groove binding molecules as well as intercalating and non-intercalating DNA binding molecules.

3. Molecules Derived from Known DNA-binding Molecules.

The assay of the present invention facilitates the identification of different binding activities by molecules derived from known DNA-binding molecules. An example of this would be to identify and test derivatives of anti-cancer drugs that have DNA-binding activity and then test for anti-cancer activity through, for example, a battery of assays performed by the National Cancer Institute (Bethesda MD). Further, the assay of the present invention can be used to test derivatives of known anti-cancer agents to examine the effect of the modifications on DNA-binding activity and specificity. In this manner, the assay may reveal activities of anti-cancer agents, and derivatives of these agents, that facilitate the design of DNA-binding

molecules with therapeutic or diagnostic applications in different fields, such as antiviral or antimicrobial therapeutics. The binding-activity information for any DNA-binding molecule, obtained by application of the present assay, can lead to a better understanding of the mode of action of more effective therapeutics.

4. Secondary Assays.

As described above, the assay of the present invention is used (i) as a screening assay to detect novel DNA-binding molecules, or (ii) to determine the relative specificity and affinity of known molecules (or their derivatives). The assay may also be used in confirmatory studies or studies to elucidate the binding characteristics of DNA-binding molecules. Using the assay as a tool for secondary studies can be of significant importance to the design of novel DNA-binding molecules with altered or enhanced binding specificities and affinities.

a.) Confirmatory Studies.

The assay of the present invention can be used in competition studies to confirm and refine the original direct binding data obtained from the assay.

The primary screening assay does not provide for the direct determination of relative absolute affinities of test molecules for different test sequences. A competition method has been developed that aids in the interpretation and confirmation of the primary screening assay. The competition method also provides a means for determining the minimum difference in absolute affinities of any test sequences for a given test molecule.

Sequences of interest are tested for their ability to compete with the test oligonucleotide for binding a

test molecule of interest. In this method, DNA molecules that contain sequences that are high affinity binding sites for the DNA-binding test molecule compete effectively with the test oligonucleotide for the binding of the test molecule. DNA molecules that contain sequences that are low affinity binding sites for the test molecules are ineffective competitors. In effect, the fold-difference in concentration required between a high affinity competitor DNA and a low affinity competitor DNA, where the competitor is required to compete with the test oligonucleotide for the binding of the DNA-binding test molecule, should be proportional to the difference in affinity between the two competitor DNA molecules.

Any test oligonucleotide may be used in the competition study. However, in practice, since most secondary screening will be used to examine the putative high affinity binding test sequences, the secondary competition assay is typically used to test a competitor oligonucleotide which is a putative high affinity test sequence.

In the competition assay, the assay conditions are essentially the same as the conditions used in the primary screening assay. The assay components are mixed, with the exception of the DNA. The mixture includes protein, buffer and the DNA-binding test molecule (control samples lack the test molecule). A test oligonucleotide is labeled (for example, using a radioisotope, although any of the described capture/detection systems should be effective in the competition study). The DNA sample, including the radiolabeled test oligonucleotide and unlabelled competitor DNA is added to the assay mixture. Typically, the competitor DNA of interest is added to different reactions over a range of competitor concentrations. Two controls are

commonly run: (i) no DNA binding test molecule added; and (ii) test DNA but no competitor DNA added.

The reactions are incubated for the desired time and the DNA:protein complexes separated from free DNA (i.e., DNA not associated with protein) by passing the mixture through nitrocellulose. Other capture systems, such as the biotin/streptavidin system discussed in Section V, are also effective. The amount of radio-labeled test oligonucleotide bound by protein (i.e., bound to the filter) is indicative of the effect of the competitor.

One example of a competition assay is as follows. A test oligonucleotide containing the test sequence TTAC ranks as a high affinity binding site for a test molecule. The TTAC test oligonucleotide is radiolabeled and mixed with non-radiolabeled competitor DNAs that contain, for example, a putative high affinity binding site (the same site, TTAC, is one example) or a putative low affinity binding site (e.g., CCCC). In the absence of any competing nonlabeled DNA or DNA-binding test molecule, the amount of radiolabeled DNA:protein complex observed (called $r\%$) is arbitrarily established as 100%. The concentration of the protein used in this experiment is high enough to bind most of the radiolabelled test oligonucleotide in the absence of test molecules or competing DNA molecules (this is essentially the same concentration as used in the primary screening assay).

The test molecule is added to the reaction at a concentration sufficient to markedly reduce $r\%$, the amount of observed DNA:protein complex. The greater the reduction in signal, the more easily competition is observed. The amount of competitor DNA needed to observe competition is proportional to the amount of DNA-binding test molecule used; therefore, the amount

of test molecule used should be sufficient to reduce $r\%$ to between approximately 10% to 70%. The effect of an effective competitor, such as TTAC, is to cause $r\%$ to rise towards 100%.

5 The competition for test molecule binding is between the non-labeled competitor DNA and the radiolabeled test oligonucleotide. As the competitor DNA concentration increases, the test molecule binds to the competitor DNA and is effectively removed from
10 solution. Accordingly, the test molecule is no longer able to block the binding of the protein to the radiolabeled oligonucleotide. A less effective competitor, typically a competitor DNA with low affinity for the test molecule, will compete less
15 effectively for the DNA-binding test molecule, even at substantially higher concentrations than the high affinity competitor. A completely ineffective competitor, i.e., one that did not bind the test molecule, would not cause the $r\%$ value to change, even at high
20 concentrations of the competitor DNA.

When a competitor DNA has some affinity for the test molecule, competition ($r\%$ rising towards 100%) would be observed at some competitor DNA concentration. The difference in concentration between two competing
25 DNA sequences to achieve an equivalent $r\%$ (e.g., 90%) should reflect the relative difference in absolute affinity between the two competitor DNA molecules. For example, if 5 μM TTAC is required to achieve a change in $r\%$ from 50% to 90% in the presence of a test mole-
30 cule and 200 μM CCCC is required to achieve the same change in $r\%$, then the fold difference in affinity between TTAC and CCCC for the test molecule is $200/5 = 40$ -fold.

In the context of screening distamycin with all
35 possible 256 bp test sequences (Example 10), the

confirmatory assay can be used (i) to confirm the rankings observed in the assay, (ii) to refine the rankings among the 5-10 highest ranked binders (which show no statistical difference in rank with data from 4 experiments), and (iii) to resolve perceived discrepancies in the assay data. All of these goals may be accomplished using a competition experiment which determines the relative ability of test sequences to compete for the binding of distamycin.

The perceived discrepancy in the distamycin experiment is as follows: test oligonucleotides scored poorly in the assay which were complementary to most of the top-ranking test sequence oligonucleotides (Examples 10 and 11). This result was unexpected since it is unlikely that the affinity of distamycin for binding a test site depends on the orientation of the screening site to the test site. More likely, the assay detects the binding of distamycin when the molecule is bound to the test oligonucleotide in one orientation, but fails to detect the binding of distamycin when the test sequence is in the other orientation. A competition study will resolve this question, since the binding of distamycin to a competitor sequence will be orientation-independent; the competition does not depend on the mechanism of the assay.

For the competition experiment, the assay may be performed under any conditions suitable for the detection of drug binding. When these conditions are established, different competitor DNAs are added to the assay system to determine their relative ability to compete for drug binding with the radiolabeled test oligonucleotide in the assay system.

The competitor DNAs may be any sequence of interest. Several classes of DNA may be tested as competitor molecules including, but not limited to, the

following: genomic DNAs, synthetic DNAs (e.g., poly(dA), poly(dI-dC), and other DNA polymers), test oligonucleotides of varying sequences, or any molecule of interest that is thought to compete for distamycin binding.

When using the competition assay to verify the results of a 256 oligonucleotide panel screen (like Example 10), the following criteria are useful for selecting the competitor test oligonucleotides:

(i) sequences that rank high in the assay but which do not have relative binding affinities with differences that are statistically significant from each other, in order to determine their relative affinity with greater precision;

(ii) sequences that are purported by other techniques (e.g., footprinting or transcriptional block analysis) to be high affinity binding sites, in order to compare the results of those techniques with the screening assay results;

(iii) sequences that are complementary to test sequences that rank high in the assay, in order to determine whether these test sequences are false negatives; and

(iv) sequences of any rank in the assay, in order to confirm the assay results.

Several methods may be used to perform the competition study as long as the relative affinities of the competing DNA molecules are detectable. One such method is described in Example 14. In this example, the concentration of the assay components (drug, protein, and DNA) is held constant relative to those used in the original screening assay, but the molar ratio of the test oligonucleotide to the competitor oligonucleotides is varied.

Another method for performing a competition assay is to hold the concentrations of protein, drug and initial amount of test oligonucleotide constant, then add a variable concentration of competitor DNA. In this design, the protein and drug concentration must be sufficiently high to allow the addition of further competitor DNA without i) decreasing the amount of DNA:protein complex in the absence of drug to a level that is unsuitable for detection of DNA:protein complex, and ii) increasing the amount of DNA:protein complex in the presence of drug to a level that is unsuitable for the detection of drug binding. The window between detectable DNA:protein complex and detectable effect of the drug must be wide enough to determine differences among competitor DNAs.

In any competition method, it is important that the relative concentrations of the competing DNA molecules are accurately determined. One method for accomplishing accurate determination of the relative concentrations of the DNA molecules is to tracer-label competitor molecules to a low specific activity with a common radiolabeled primer (Example 14). In this manner, the competitor molecules have the same specific activity, but are not sufficiently radioactive (200-fold less than the test oligonucleotide) to contribute to the overall radioactivity in the assay.

b.) Secondary Studies to Elucidate Binding Characteristics. The studies outlined in Section VI.B.4.a describe methods of determining some of the binding processes of distamycin A. The assay of the present invention may also be used to explore mechanistic questions about distamycin binding.

For example, several of the complements of the putative high affinity binding sites for distamycin

have low scores in the assay. As described above, this may imply directionality in binding. The results may also imply that the test sites are not equal with respect to the effect exerted on UL9-COOH binding.

5 Oligonucleotides can be designed to test the hypothesis of directionality.

The basic test oligonucleotide has the structure presented in Figure 27A (SEQ ID NO:621). In one scenario, the score in the binding assay is high, i.e.,
10 the greatest effect of distamycin, when the test sequences is XYZZ (Figure 27A, with the base X complementary to the base Y and the base Q complementary to the base Z), and the complement (Figure 27B; SEQ ID NO:622) scores low. These results imply that the test sites
15 are not equivalent with respect to their effect on UL9, otherwise the right side would have the effect in one oligonucleotide and the left site would have the effect in the other. These results further suggest that the effect of distamycin is directional. The only assumption
20 is that distamycin should bind with the same affinity to the XYZZ/QQXY sequence (Figures 27A and 27B) regardless of its position or orientation in the oligonucleotide. Since the scores are derived at equilibrium, this is likely to be the case.

25 To test the hypothesis that one site is effective in the assay, oligonucleotides may be designed that have the UL9 site inverted with respect to the test sites (Figures 27C and 27D; SEQ ID NO:623 and SEQ ID NO:624, respectively). If only one site is active with
30 respect to UL9 and if the Figure 27A oligo was most effective in binding distamycin, then the oligo C should be less active in the assay than oligo D; in other words, flipping the UL9 site will result in QQXY ranking high, XYZZ ranking low.

Finally, to determine the "direction" of distamycin binding, mix test sequences and invert the binding site as shown in the four oligonucleotides presented in Figures 27E, 27F, 27G and 27H. Alternatively, one test site or the other could be deleted from the test oligonucleotide.

This type of analysis provides an example of the usefulness in the assay in determining binding properties of DNA-binding drugs.

10

c.) Restriction Endonucleases as Indicator Proteins in the Assay. Other DNA:protein interactions that are useful as screening sequences and their cognate binding proteins (indicator proteins) are restriction enzymes. Such secondary screening assays are performed using the same criteria to establish conditions for the primary screening assay (described in Example 4). The assay conditions can be varied to accommodate different DNA:protein interactions, as long as the assay system follows the functional criteria discussed above (Section I).

One limitation of using restriction endonucleases in the method of the present invention is that the assay buffer should not contain divalent cations. In the absence of divalent cations, the enzymes will bind the appropriate recognition sequence, but not cleave the DNA. In the presence of divalent cations, the test oligonucleotide can be cleaved at or near the protein binding site.

By using different indicator proteins, a different recognition sequence can be used to flank the test site. This variation allows the resolution of questions regarding the potential binding of a test molecule to a site internal to any single screening sequence. For example, the assay system is used where

the UL9 protein and its recognition sequence are used as the indicator protein:screening sequence interaction. In this system, if the highest affinity binding site for a test molecule is TTAC, then several test sequences may be predicted to rank high in the assay system: several of these test sequences are presented in Figure 31. In Figure 31, the test site is shown in bold, the potential binding site for the test molecule is shown underlined.

One test oligonucleotide on which the DNA-binding test molecule would be predicted to have a high level of effect is the oligonucleotide containing the test site, TTAC (Figure 31). However, since the UL9 recognition sequence contains the sequence TT, flanking the test site, several other test oligonucleotides might also be expected to have high activity in the assay (see Figure 31).

By using a different DNA:protein interaction as the indicator system in a secondary screening assay, the "false positives" shown for TACN and ACNN (shown in Figure 31) can be identified. The recognition sequence for the protein in a secondary screening assay simply needs to have a different screening sequence in the region flanking the test site than the UL9 screening sequence.

Restriction endonucleases provide an entire class of different DNA:protein interactions with a wide array of available sequences that can be used in this manner. For example, *SmaI* recognizes the sequence 5'-CCCGGG-3'. Using the *SmaI*:DNA interaction and the same test sequences presented in Figure 31, the resulting test oligonucleotides would have the test sequences presented in Figure 32. As can be seen from a comparison of Figures 31 and 32, changing the screening sequence from the UL9-binding sequence to the *SmaI*-binding sequence

eliminates the potential test molecule binding sites internal to the screening sequence (e.g., compare TACN and ACNN in the figures).

The use of different DNA-binding proteins as indicator proteins in the assay is also applicable to the PCR-based test oligonucleotide selection technology (Section III).

10 5. Generation of Binding Data and Refinement of Molecular Modeling Systems.

The assay of the present invention generates data which can be applied to the refinement of molecular modeling systems that address DNA structural analysis: the data is also useful in the design and/or refinement of DNA-binding drugs. Traditionally, mass screening has been the only reasonable method for discovering new drugs. Modern rational drug design seeks to minimize laboratory screening. However, *ab initio* rational drug design is difficult at this time given (i) insufficiencies in the underlying theories used for *de novo* design, and (ii) the computational intensity which accompanies such design approaches.

The *ab initio* approach requires calculations from first principles by quantum mechanics: such an approach is expensive and time-consuming. The introduction of data concerning the relative binding affinities of one or more DNA-binding molecules to all 256 four base pair DNA sequences allows the development, via molecular modeling, of *ad hoc* protocols for DNA structural analysis and subsequent DNA-binding drug design. The accumulation of data for the DNA sequences to which small molecules bind is likely to result in more accurate, less expensive molecular modeling programs for the analysis of DNA.

The screening capacity of the assay of the present invention is much greater than screening a single DNA sequence with an individual cognate DNA-binding protein. Direct competition assays involving individual receptor:ligand complexes (e.g., a specific DNA:protein complex) are most commonly used for mass screening efforts. Each such assay requires the identification, isolation, purification, and production of the assay components. In particular, a suitable DNA:protein interactions must be identified for each selected screening sequence. Using the assay of the present invention, libraries of synthetic chemicals or biological molecules can be screened to detect molecules that have preferential binding to virtually any specified DNA sequence -- all using a single assay system. When employing the assay of the present invention, secondary screens involving the specific DNA:protein interaction may not be necessary, since inhibitory molecules detected in the assay may be tested directly on a biological system: for example, the ability to disrupt viral replication in a tissue culture or animal model.

6. The Design of New DNA-Binding Heteropolymers Comprised of Subunits Directed to Different DNA Sequences.

The assay of the present invention will facilitate the predictive abilities of molecular modeling systems in two ways. First, *ad hoc* methods of structural prediction will be improved. Second, by employing pattern matching schemes, the comparison of sequences having similar or different affinities for a given set of DNA-binding molecules should empirically reveal sets of sequences that have similar structures (see Section VI.D, Using a Test Matrix). Molecular modeling programs are "trained" using the information

concerning DNA-binding molecules and their preferred binding sequences. With this information coupled to the predicative power of molecular modeling programs, the design of DNA-binding molecules (subunits) that
5 could be covalently linked becomes feasible.

These molecular subunits would be directed at defined sections of DNA. For example, a subunit would be designed for each possible DNA unit. For example, if single bases were the binding target of the sub-
10 units, then four subunits would be required, one to correspond to each base pair. These subunits could then be linked together to form a DNA-binding polymer, where the DNA binding preference of the polymer corresponds to the sequence binding preferences of the
15 subunits in the particular order in which the subunits are assembled.

Another example of such a polymer is using subunits whose binding was directed at two base sections of DNA. In this case, $4^2 = 16$ subunits would
20 be used, each subunit having a binding affinity for a specific two base pair sequence (e.g., AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG, TT). If the polymers were to be comprised of subunits targeted to 3 base pair sections of DNA, then $4^3 = 64$ subunits
25 would be prepared. The design of such molecular subunits is dependent upon the establishment of a refined database using empirical data derived by the method of the present invention, as described in Section VI.B.

30

C. Sequences Targeted by the Assay.

The DNA:protein assay of the present invention has been designed to screen for compounds that bind a full range of DNA sequences that vary in length as well as
35 complexity. Sequence-specific DNA-binding molecules

discovered by the assay have potential usefulness as either molecular reagents, therapeutics, or therapeutic precursors. Sequence-specific DNA-binding molecules are potentially powerful therapeutics for essentially
5 any disease or condition that in some way involves DNA. Examples of test sequences for the assay include: a) binding sequences of factors involved in the maintenance or propagation of infectious agents, especially viruses, bacteria, yeast and other fungi, b) sequences
10 causing the inappropriate expression of certain cellular genes, and c) sequences involved in the replication of rapidly growing cells. Furthermore, gene expression or replication need not necessarily be disrupted by blocking the binding of specific proteins.
15 Specific sequences within protein-coding regions of genes (e.g., oncogenes) are equally valid test sequences since the binding of small molecules to these sequences is likely to perturb the transcription and/or replication of the region. Finally, any molecules that
20 bind DNA with some sequence specificity, that is, not just to one particular test sequence, may be still be useful as anti-cancer agents. Several small molecules with some sequence preference are already in use as anticancer therapeutics. Molecules identified by the
25 present assay may be particularly valuable as lead compounds for the development of congeners having either different specificity or different affinity.

One advantage of the present invention is that the assay is capable of screening for binding activity
30 directed against any DNA sequence. Such sequences can be medically significant target sequences scrambled or randomly generated DNA sequences, or well-defined, ordered sets of DNA sequences. Other sets could be used for screening for molecules demonstrating sequence
35 preferential binding (like Doxorubicin) to determine

the sequences with highest binding affinity and/or to determine the relative affinities between a large number of different sequences. There is usefulness in taking either approach for detecting and/or designing new therapeutic agents. Section VI.C.3, "Theoretical Considerations for Choosing Target Sequences", outlines the theoretical considerations for choosing DNA target sites in a biological system.

10 1. Medically Significant Target Sequences.

 Few effective viral therapeutics are currently available; yet several potential target sequences for antiviral DNA-binding drugs have been well-characterized. Furthermore, with the accumulation of sequence data on all biological systems, including viral genomes, cellular genomes, pathogen genomes (bacteria, fungi, eukaryotic parasites, etc.), the number of target sites for DNA-binding drugs will increase greatly in the future.

20 There are numerous methods for identifying medically significant target sequences for DNA-binding drugs, including, but not limited to, the following. First, medically significant target sequences are found in pathogens of the biological kingdoms, for example in genetic sequences that are key to biochemical pathways or physiological processes. Second, a target is identified, such as (i) a pathogen involved in an infectious disease, or (ii) a biochemical pathway or physiological process of a noninfectious disease, genetic condition, or other biological process. Then specific genes important for the survival of the pathogen or modulation of the endogenous pathway involved in the target system are identified. Third, specific target sequences are identified that affect

the expression or activity of a DNA molecule, such as genes or sites involved in replication.

There are numerous pathogens that are potential targets for DNA-binding drugs designed using the methods described in this application. Table I lists a number of potential target pathogens.

Table I: Pathogens	
VIRUSES	
10	Retroviruses Human HIV I ,II HTLV I, II
15	Animal SIV STLV I FELV FIV BLV BIV (Bovine immunodeficiency virus) Lentiviruses Avian reticuloendotheliosis virus
20	
25	Animal - continued SIV STLV I FELV FIV BLV BIV (Bovine immunodeficiency virus) Lentiviruses Avian reticuloendotheliosis virus Avian sarcoma and leukosis viruses Caprine arthritis-encephalitis Equine infectious anemia virus Maedi/visna of sheep MMTV (mouse mammary tumor virus) Progressive pneumonia virus of sheep
30	
35	
40	Herpesviridae Human EBV CMV

	HSV I, II
	VZV
	HH6
5	Cercopthecine Herpes Virus (B Virus)
	Old world monkeys with infection into humans.
	Animal
	Bovine Mammillitis virus
	Equine Herpes virus
10	Equine coital exanthema virus
	Equine rhinopneumonitis virus
	Infectious bovine rhinotracheitis virus
	Marek's disease virus of fowl
	Turkey herpesvirus
15	Hepadnaviruses
	Human
	HBV/HDV
	Animal
20	Duck Hepatitis
	Woodchucks
	Squirrels
	Poxviridae
	Human
25	Orf virus
	Cow Pox
	Variola virus
	Vaccinia
	Small Pox
	Pseudocowpox
30	Poxviridae - continued
	Animal
	Bovine papular stomatitis virus
	Cowpox virus
35	Ectromelia virus (mouse pox)
	Fibroma viruses of rabbits/squirrels
	Fowlpox
	Lumpy skin disease of cattle virus
	Myxoma
	Pseudocowpox virus
40	Sheep pox virus
	Swine pox

5	Papovaviridae Human BK virus SV-40 JC virus Human Papillomaviruses 1-58 (see list Fields) Animal Lymphotropic papovavirus (LPV) Monkey Bovine papillomavirus Shope papillomavirus
10	
15	Adenoviridae Human Adenoviruses 1-4 Animal Canine adenoviruses 2
20	Parvoviridae Human AAV (Adeno Associated Virus) B19 (human)
25	Animal FPV (Feline parvovirus) PPV (Porcine parvovirus) ADV (Aleutian disease, mink) Bovine Parvovirus Canine Parvovirus Feline panleukopenia virus Minute virus of mice Mink enteritis virus
30	
	BACTERIA
35	Streptococcus <i>pneumonia</i> <i>bovis</i>
40	Group A Streptococci Agents responsible for: Streptococcal pharyngitis Cervical adenitis Otitis media Mastoiditis Peritonsillar abscesses Meningitis Peritonitis

	Pneumonia
	Acute glomerulonephritis
	Rheumatic fever
	Erythema nodosum
5	<i>Staphylococcus</i>
	<i>aureus</i>
	<i>epidermidis</i>
	<i>saprophyticus</i>
	<i>cohnii</i>
10	<i>haemolyticus</i>
	<i>xylosus</i>
	<i>warneri</i>
15	<i>capitis</i>
	<i>hominis</i>
	<i>simulans</i>
	<i>saccharolyticus</i>
	<i>auricularis</i>
20	Agents responsible for:
	Furuncles
	Carbuncles
	Osteomyelitis
	Deep tissue abscesses
	Wound infections
25	Pneumonia
	Empyema
	Pericarditis
	Endocarditis
	Meningitis
30	Purulent arthritis
	Enterotoxin in food poisoning
	<i>Branhamella catarrhalis</i>
	<i>Neisseria</i>
35	<i>gonorrhoea</i>
	<i>lactamica</i>
	<i>sicca</i>
	<i>subflava</i>
	<i>mucosa</i>
40	<i>Neisseria</i> - continued
	<i>flavescens</i>
	<i>cinerea</i>
	<i>elongata</i>
	<i>canis</i>
	<i>meningitides</i>

	Enteric Bacilli and Similar Gram-Negative Bacteria
5	<i>Escherichia</i>
	<i>Proteus</i>
	<i>Klebsiella</i>
	<i>Pseudomonas aeruginosa</i>
	<i>Enterobacter</i>
	<i>Citrobacter</i>
10	<i>Proteus</i>
	<i>Providencia</i>
	<i>Bacteroides</i>
	<i>Serratia</i>
15	<i>Pseudomonas</i> (not <i>aeruginosa</i>)
	<i>Acinetobacter</i>
	<i>Salmonella</i>
	<i>Shigella</i>
	<i>Aeromonas</i>
20	<i>Moraxella</i>
	<i>Edwardsiella</i>
	<i>Ewingella</i>
	<i>Hafnia</i>
	<i>Kluyvera</i>
	<i>Morganella</i>
25	<i>Plesiomonas</i>
	<i>Pseudomonas</i>
	<i>aeruginosa</i>
	<i>putida</i>
30	<i>pseudomallei</i>
	<i>mallei</i>
	<i>Haemophilus</i>
	<i>ducreyi</i>
	<i>influenzae</i>
	<i>parainfluenzae</i>
35	<i>Bordetella pertussis</i>
	<i>Yersinia</i>
	<i>pestis</i> (plague)
	<i>pseudotuberculosis</i>
	<i>enterocolitica</i>
40	<i>Francisella tularensis</i>
	<i>Pasteurella multocida</i>
	<i>Vibrio</i>
	<i>cholerae</i>
	<i>parhaemolyticus</i>

100

	<i>fluvialis</i>
	<i>furnissii</i>
	<i>mimicus</i>
5	<i>Brucella</i> <i>melitensis</i> <i>abortus</i> <i>suis</i> <i>canis</i>
	<i>Bartonella bacilliformis</i>
10	<i>Gardnerella vaginalis</i>
15	<i>Borrelia</i> <i>recurrentis</i> <i>hermsii</i> <i>duttoni</i> <i>crocidurae</i> <i>burgdorferi</i> (Lyme disease)
20	<i>Bacillus</i> <i>anthracis</i> <i>cereus</i> <i>megaterium</i> <i>subtilis</i> <i>sphaericus</i> <i>circulans</i> <i>brevis</i>
25	<i>lentiformis</i> <i>macerans</i> <i>pumilus</i> <i>thuringiensis</i> <i>larvae</i>
30	<i>lentimorbus</i> <i>popilliae</i>
	<i>Streptobacillus moniliformis</i> (rat bite fever)
	<i>Spirillum minus</i> (rat bite fever)
	<i>Rothia dentocariosa</i>
35	<i>Kurthia</i>
	<i>Clostridium</i> <i>botulinum</i> <i>nouyi</i> <i>bifermentans</i>

5	<i>Clostridium - continued</i> <i>histolyticum</i> <i>ramosum</i> <i>tetani</i> <i>perfringens</i> <i>novyi</i> <i>septicum</i>
10	<i>Campylobacter</i> <i>jejuni</i> <i>fetus</i> <i>hyintestinalis</i> <i>fennelliae</i> <i>cinaedi</i>
15	<i>Corynebacterium</i> <i>ulcerans</i> <i>pseudotuberculosis</i> <i>JK</i> <i>diphtheriae</i>
20	<i>Legionella</i> <i>pneumophila</i> <i>bosemanii</i> <i>micdadie</i> <i>bozenamii</i> <i>feleii</i>
25	<i>many others</i>
30	<i>Mycobacterium</i> <i>tuberculosis</i> <i>africanum</i> <i>bovis</i> <i>leprae</i> <i>avium complex</i> <i>kansasii</i> <i>fortuitum complex</i> <i>scrofulaceum</i>
35	<i>marinum</i> <i>ulcerans</i>
	<i>Actinomyces</i>
	<i>Bacteroides</i> <i>fragilidis</i>
40	<i>Fusobacterium</i> <i>necrophorum</i> <i>nucleatum</i>
	<i>Peptostreptococcus</i>
	<i>Arachnia</i>

	<i>Bifidobacterium</i>
	<i>Propionibacterium</i>
	<i>Nocardia</i>
	<i>Treponema pallidum</i> (syphilis)
5	<i>Rickettsiae</i>
	Typhus
	<i>R. prowazeki</i> (epidemic)
	<i>R. prowazeki</i> (Brill's disease)
	<i>R. typhi</i> (endemic)
10	Spotted fever
	<i>R. rickettsi</i>
	<i>R. sibiricus</i>
	<i>R. conorii</i>
	<i>R. australis</i>
15	<i>R. akari</i>
	Scrub typhus
	<i>R. tsutsugamushi</i>
	Q fever
	<i>Coxiella burnetii</i>
20	Trench fever
	<i>Rochalimaea quintana</i>
	<i>Chlamydiae</i>
	<i>C. trachomatis</i>
25	(blindness, pelvic inflammatory disease, LGV)
	<i>Mycoplasma</i>
	<i>pneumoniae</i>
	<i>Ureaplasma urealyticum</i>
	<i>Cardiobacterium hominis</i>
30	<i>Actinobacillus actinomycetemcomitans</i>
	<i>Kingella</i>
	<i>Capnocytophaga</i>
	<i>Pasteurella multocida</i>
	<i>Leptospira interrogans</i>
35	<i>Listeria monocytogenes</i>
	<i>Erysipelothrix rhusiopathiae</i>
	<i>Streptobacillus moniliformis</i>
	<i>Calymatobacterium granulomatis</i>
	<i>Bartonella bacilliformis</i>

	<i>Francisella tularensis</i>
	<i>Salmonella typhi</i>
	FUNGAL
5	<i>Actinomyces</i> <i>israelii</i> <i>naeslundii</i> <i>viscosus</i> <i>odontolyticus</i> <i>meyeri</i> 10 <i>pyogenes</i>
	<i>Cryptococcus neoformans</i>
	<i>Blastomyces dermatitidis</i>
	<i>Histoplasma capsulatum</i>
	<i>Coccidioides immitis</i>
15	<i>Paracoccidioides brasiliensis</i>
	<i>Candida</i> <i>albicans</i> <i>tropicalis</i> (<i>Torulopsis</i>) <i>glabrata</i> 20 <i>parapsilosis</i>
	<i>Aspergillus</i> <i>fumigatus</i> <i>flavus</i> <i>niger</i> 25 <i>terreus</i>
	<i>Rhinosporidiosis seeberi</i>
	<i>Phycomycetes</i>
	<i>Sporothrix schenickii</i>
	<i>Mucorales</i>
30	<i>Entomophthorales</i>
	Agents of <i>Chromoblastomycosis</i>
	<i>Microsporum</i> <i>M. audouinii</i> (ring worm) <i>M. canis</i> 35 <i>M. gypseum</i>

5	<i>Trichophyton</i> <i>T. schoenleinii</i> (favus-ringworm) <i>T. violaceum</i> (hair) <i>T. tonsurans</i> (hair) <i>T. mentagrophytes</i> (athlete's foot) <i>T. rubrum</i> (athlete's foot)
	<i>Malassezia furfur</i>
10	<i>Cladosporium</i> <i>werneckii</i> <i>carrioni</i>
	<i>Fonsecaea</i> <i>pedrosoi</i> <i>compacta</i>
	<i>Phialophora verrucosa</i>
15	<i>Rhinocladiella aquaspersa</i>
	<i>Trichosporon cutaneum</i>
	<i>Piedraia hortai</i>
	<i>Ascomycota</i>
	<i>Basidiomycota</i>
20	<i>Deuteromycota</i>
	<i>Norcardia</i> <i>brasiliensis</i> <i>caviae</i> <i>asteroides</i>
25	PARASITIC PATHOGENS
30	<i>Plasmodium</i> (malaria) <i>falciparum</i> <i>vivax</i> <i>ovale</i> <i>malariae</i>
35	<i>Schistosoma</i> <i>japonicum</i> <i>mansoni</i> <i>haematobium</i> <i>intercalatum</i> <i>mekongi</i>

5	<i>Trypanosoma</i> <i>brucei gambiense</i> <i>brucei rhodesiense</i> <i>evansi</i> <i>cruzi</i> <i>equiperdum</i> <i>congolense</i>
	<i>Entamoeba histolytica</i>
	<i>Naegleria fowleri</i>
10	<i>Acanthamoeba</i> <i>astronyxis</i> <i>castellanii</i> <i>culbertsoni</i> <i>hatchetti</i> <i>palestinensis</i> <i>polyphaga</i> <i>rhyusodes</i>
15	
20	<i>Leishmania</i> <i>dovonani</i> <i>infantum</i> <i>chagasi</i> <i>topica</i> <i>major</i> <i>aethiopica</i> <i>mexicana</i> <i>braziliensis</i> <i>peruviana</i>
25	
	<i>Pneumocystis carinii</i> (interstitial pneumonia)
30	<i>Babesia</i> (tick born hemoprotzoan) <i>microti</i> <i>divergens</i>
	<i>Giardia lamblia</i>
35	<i>Trichomonas</i> (venereal disease) <i>vaginalis</i> <i>hominis</i> <i>tenax</i>
	<i>Cryptosporidium parvum</i> (intestinal protozoan)
	<i>Isopora belli</i> (dysentery)
	<i>Balantidium coli</i> (protozoon induced dysentery)
40	<i>Dientamoeba fragilis</i>
	<i>Blastocystis hominis</i>

	<i>Trichinella spiralis</i> (parasitic nematode)
	<i>Wuchereria bancrofti</i> (lymphatic filariasis)
5	<i>Brugia</i> (lymphatic filariasis) <i>malayi</i> <i>timori</i>
	<i>Loa loa</i> (eye worm)
	<i>Onchocerca volvulus</i>
10	<i>Mansonella</i> <i>perstans</i> <i>ozzardi</i> <i>streptocerca</i>
	<i>Dirofilaria immitis</i>
15	<i>Angiostrongylus cantonensis</i> <i>costaricensis</i> <i>malayensis</i> <i>mackerrasae</i>
	<i>Anisakis</i> (nematode) <i>simplex</i> <i>typica</i>
20	<i>Pseudoterranova decipiens</i>
	<i>Gnathostoma spinigerum</i>
	<i>Dracunculus medinensis</i> (filarial parasite, guinea worm)
	<i>Trichuris trichiura</i> (whip worm)
25	<i>Ascaris lumbricoides</i> (nematode)
	<i>Toxocara canis</i> (nematode round worms)
	<i>Necator americanus</i> (heart worm)
30	<i>Ancylostoma</i> (hook worm) <i>duodenale</i> <i>ceylanicum</i> <i>americanus</i>
	members of the species <i>Trichostrongylus</i>
35	<i>Strongyloides</i> (intestinal nematode) <i>stercoralis</i> <i>fuelleborni</i>
	<i>Capillaria philippinensis</i> (intestinal nematode)

	Various species of <i>Paragonimus</i> (lung fluke disease)
	Various species of <i>Micorsporida</i>
	<i>Clonorchis sinensis</i> (liver fluke)
5	<i>Fasciola</i> (trematode, intestinal worm) <i>hepatica</i> <i>gigantica</i>
	<i>Fasciolopsis buski</i>
	<i>Heterophyes heterophyes</i>
10	<i>Metagonimus yakagawa</i>
	<i>Taenia</i> <i>saginata</i> (beef tapeworm) <i>solium</i> (pork tapeworm)
15	<i>Hymenolepis</i> (dwarf tapeworm) <i>nana</i> <i>nana fraterna</i> <i>diminuta</i>
	<i>Dipylidium caninum</i> (tapeworm of dogs and cats)
20	<i>Diphyllbothrium</i> (fish tapeworms) <i>lantum</i> <i>dalliae</i> <i>nihonkaiense</i> <i>pacificum</i>
25	<i>Echinococcus</i> (tape worm with cysts) <i>granulosus</i> <i>multilocularis</i> <i>vogeli</i>
30	<i>Enterobius vermicularis</i> (Pin worm)

In addition to pathogens, many non-infectious diseases may be controlled at the level of DNA. These diseases are therefore potential candidates for treatment with DNA-binding therapeutics that are discovered or designed using the methods described in this application. Table II lists a number of potential non-infectious diseases that may be targeted for treatment using DNA-binding molecules.

Table II: Noninfectious Diseases	
CANCER	
5	Lung Adenocarcinoma Squamous cell Small cell
	Breast carcinoma
10	Ovarian Serous tumors Mucinous tumors Endometrioid carcinoma
	Endometrial carcinoma
	Colon carcinoma
	Malignant Melanoma
15	Prostate carcinoma
	Lymphoma Hodgkins Non-Hodgkin's
20	Leukemia Chronic Myelogenous Acute Myelogenous Chronic Lymphocytic Acute Lymphocytic
	Cervical carcinoma
25	Seminoma
	Multiple Myeloma
	Bladder carcinoma
	Pancreatic carcinoma
	Stomach carcinoma
30	Thyroid Papillary adenocarcinoma Follicular carcinoma Medullary carcinoma
	Oral & Pharyngeal carcinomas
35	Laryngeal carcinoma
	Bladder carcinoma

	Renal cell carcinoma
	Hepatocellular carcinoma
	Glioblastoma
	Astrocytoma
5	Meningioma
	Osteosarcoma
	Pheochromocytoma
	CARDIOVASCULAR DISEASES
10	Hypertension Essential Malignant
	Acute Myocardial Infarction
15	Stroke Ischemic Hemorrhagic
	Angina Pectoris
	Unstable angina
	Congestive Heart Failure
	Supraventricular arrhythmias
20	Ventricular arrhythmias
	Deep Venous Thrombosis
	Pulmonary Embolism
	Hypercholesterolemia
	Cardiomyopathy
25	Hypertriglyceridemia
	RESPIRATORY DISORDERS
	Allergic rhinitis
	Asthma
	Emphysema
30	Chronic bronchitis
	Cystic Fibrosis
	Pneumoconiosis

	Respiratory distress syndrome
	Idiopathic pulmonary fibrosis
	Primary pulmonary hypertension
	GASTROINTESTINAL DISORDERS
5	Peptic ulcers
	Cholelithiasis
	Ulcerative colitis
	Crohn's disease
	Irritable Bowel Syndrome
10	Gastritis
	Gilbert's syndrome
	Nausea
	ENDOCRINE/METABOLIC DISORDERS
	Diabetes mellitus type I
15	Diabetes mellitus type II
	Diabetes insipidus
	Hypothyroidism
	Hyperthyroidism
	Gout
20	Wilson's disease
	Addison's disease
	Cushing's syndrome
	Acromegaly
	Dwarfism
25	Prolactinemia
	Morbid obesity
	Hyperparathyroidism
	Hypoparathyroidism
30	Osteomalacia

RHEUMATOLOGY/IMMUNOLOGY DISORDERS	
	Transplant rejection
	Systemic lupus erythematosus
	Rheumatoid arthritis
5	Temporal Arteritis
	Amyloidosis
	Sarcoidosis
	Sjogren's Syndrome
	Scleroderma
10	Ankylosing spondylitis
	Polymyositis
	Reiter's Syndrome
	Polyarteritis nodosa
	Kawasaki's disease
15	HEMATOLOGIC DISORDERS
	Anemia
	Sickle cell
	Sideroblastic
	Hereditary spherocytosis
20	Aplastic
	Autoimmune hemolytic anemia
	Thalassemia
	Disseminated intravascular coagulation
	Polycythemia vera
25	Thrombocytopenia
	Thrombotic thrombocytopenic purpura
	Idiopathic thrombocytopenic purpura
	Hemophilia
	von Willebrand's disease
30	Neutropenia
	Post-chemotherapy
	Post-radiation

	NEUROLOGIC DISORDERS
	Alzheimer's disease
	Parkinson's disease
	Myasthenia gravis
5	Multiple sclerosis
	Amyotrophic lateral sclerosis
	Epilepsy
10	Headaches Migraine Cluster Tension
	Guillain-Barre syndrome
	Pain (post-op, trauma)
	Vertigo
15	PSYCHIATRIC DISORDERS
	Anxiety
	Depression
	Schizophrenia
	Substance abuse
20	Manic-Depression
	Anorexia
	DERMATOLOGIC DISORDERS
	Acne
	Psoriasis
25	Eczema
	Contact dermatitis
	Pruritis
	OPHTHALMIC DISORDERS
	Glaucoma
30	Allergic conjunctivitis
	Macular degeneration

	MUSCULOSKELETAL DISORDERS
	Osteoporosis
	Muscular dystrophy
	Osteoarthritis
5	GENETIC DISORDERS
	Down's syndrome
	Marfan's syndrome
	Neurofibromatosis
	Tay-Sachs disease
10	Gaucher's disease
	Niemann-Pick disease
	GENITAL-URINARY DISORDERS
	Benign prostatic hypertrophy
	Polycystic kidney disease
15	Non-infectious glomerulonephritis
	Goodpasture's syndrome
	Urolithiasis
	Endometriosis
	Impotence
20	Infertility
	Fertility control
	Menopause

25 Once a disease or condition is identified as a
 potential candidate for treatment with a DNA-binding
 therapeutic, specific genes or other DNA sequences that
 are crucial for the expression of the disease associat-
 ed gene (or survival of a pathogen) are identified
 within the biochemical or physiological pathway (or the
 30 pathogen). In humans, many genes involved in important
 biological functions have been identified. Virtually

any DNA sequence is a potential target site for a DNA-binding molecule, including mRNA coding sequences, promoter sequences, origins of replication, and structural sequences, such as telomeres and centromeres. One class of sites that may be preferable are the recognition sequences for proteins that are involved in the regulation or expression of genetic material. For this reason, the promoter/regulatory regions of genes also provide potential target sites (Table III, see also Example 15).

Table III: Human Genes with Promoter Regions that are Potential Targets for DNA-Binding Molecules

*[LOCUS Names are from EMBL database ver. 33. 1992.]

LOCUS Names*	Locus Description
>HS5FDX	Human ferredoxin gene, 5' end.
>HSA1ATCA	Human macrophage alpha1-antitrypsin cap site region
>HSA1GPB1	Human gene B for alpha 1-acid glycoprotein exon 1 and 5' flank
>HSA1MBG1	Human gene for alpha-1-micro-globulin-bikunin, exons 1-5 (encoding
>HSA2MGLB1	H.sapiens gene for alpha-2 macroglobulin, exon 1
>HSACAA1	H.sapiens ACAA gene (exons 1 & 2) for peroxisomal 3-oxoacyl-CoA
>HSACCOA	Homo sapiens choline acetyltransferase gene sequence.
>HSACEB	Human angiotensin I-converting enzyme (ACE) gene, 5' flank.
>HSACHG1	Human gene fragment for the acetylcholine receptor gamma subunit
>HSACT2CK1	Human cytokine (Act-2) gene, exon 1.
>HSACTBPR	Human beta-actin gene 5'-flanking region

	>HSACTCA	Human cardiac actin gene, 5' flank.
	>HSACTSA	Human gene for vascular smooth muscle alpha-actin (ACTSA), 5'
	>HSACTSG1	Human enteric smooth muscle gamma-actin gene, exon 1.
	>HSAD12L	Human arachidonate 12-lipoxygenase gene, 5' end.
5	>HSADH1X	Human alcohol dehydrogenase alpha subunit (ADH1) gene, exon 1.
	>HSADH2X	Human alcohol dehydrogenase beta subunit (ADH2) gene, exon 1.
	>HSAFPCP	Human alpha-fetoprotein gene, complete cds.
	>HSAK1	Human cytosolic adenylate kinase (AK1) gene, complete cds.
	>HSAGAL	Human alpha-N- acetylgalactosaminidase (NAGA) gene, complete cds.
10	>HSALADG	H.sapiens ALAD gene for porphobilinogen synthase
	>HSALBENH	Human albumin gene enhancer region.
	>HSALDA1	Human aldolase A gene 5' non-coding exons
	>HSALDCG	Human aldolase C gene for fructose-1,6-bisphosphate aldolase
	>HSALDOA	Human aldolase A gene (EC 4.1.2.13)
15	>HSALDOBG	Human DNA for aldolase B transcription start region
	>HSALIFA	Human leukemia inhibitory factor (LIF) gene, complete cds.
	>HSAMINON	Human aminopeptidase N gene, complete cds.
	>HSAMY2A1	Human alpha-amylase (EC 3.2.1.1) gene AMY2A 5-flank and exon 1
	>HSAMYB01	Human amyloid-beta protein (APP) gene, exon 1. 1154
20	>HSANFG1	Human gene fragment for pronatriodilatin precursor (exons 1 and 2)

	>HSANFPRE	Human gene for atrial natriuretic factor (hANF) precursor
	>HSANFZ1	Human atrial natriuretic factor gene, complete cds.
	>HSANGG1	Human angiotensinogen gene 5' region and exon 1
	>HSANT1	Human heart/skeletal muscle ATP/ADP translocator (ANT1) gene,
5	>HSAPC3A	Human apolipoprotein CIII gene and apo AI-apo CIII intergenic
	>HSAPC3G	Human gene for apolipoprotein C-III
	>HSAPOA2	Human gene for apolipoprotein AII
	>HSAPOAIA	Human fetal gene for apolipoprotein AI precursor
	>HSAPOBPRM	Human apoB gene 5' regulatory region (apolipoprotein B)
10	>HSAPOC2G	Human apoC-II gene for preproapolipoprotein C-II
	>HSAPOCIA	Human apolipoprotein C-I (VLDL) gene, complete cds.
	>HSAPOLIDG	H.sapiens promoter region of gene for apolipoprotein D
	>HSARG1	Human arginase gene exon 1 and flanking regions (EC 3.5.3.1)
	>HSASG5E	Human argininosuccinate synthetase gene 5' end 1105
15	>HSATP1A3S	Human sodium/potassium ATPase alpha 3 subunit (ATP1 A3) gene, 5'
	>HSBSF2	Human (BSF-2/IL6) gene for B cell stimulatory factor-2
	>HSC5GN	Human C5 gene, 5' end. 650
	>HSCAII	Human gene fragment for carbonic anhydrase II (exons 1 and 2)
	>HSCALCAC	Human calcitonin/alpha-CGRP gene
20	>HSCALRT1	Human DNA for calretinin exon 1
	>HSCAPG	Human cathepsin G gene, complete cds.

	>HSCAVIII1	H.sapiens carbonic anhydrase VII (CA VII) gene, exon 1.
	>HSCBMYHC	Human gene for cardiac beta myosin heavy chain
	>HSCD3AA	Human complement C3 protein mRNA, 5' flank. >HSCD4 Human recognition/surface antigen (CD4) gene, 5' end.
	>HSCD44A	Human hyaluronate receptor (CD44) gene, exon 1.
5	>HSCFTC	Human cystic fibrosis transmembrane conductance regulator gene, 5'
	>HSCH7AHYR	Human cholesterol 7-alpha-hydroxylase (CYP7) gene, 5' end.
	>HSCHAT	Human gene for choline acetyltransferase (EC 2.3.1.6), partial
	>HSCHYMASE	Human mast cell chymase gene, complete cds.
	>HSCHYMB	Human heart chymase gene, complete cds. 3279
10	>HSCKBG	Human gene for creatine kinase B (EC 2.7.3.2)
	>HSCNP	Human C-type natriuretic peptide gene, complete cds.
	>HSCD59011	Human transmembrane protein (CD59) gene, exon 1.
	>HSCDPRO	Human myeloid specific CD11b promoter DNA.
	>HSCETP1	Human cholesteryl ester transfer protein (CETP) gene, exons 1 and
15	>HSCFTC	Human cystic fibrosis transmembrane conductance regulator gene, 5'
	>HSCOSEG	H.sapiens coseg gene for vasopressin-neurophysin precursor
	>HSCREKIN	Human creatine kinase gene, exon 1.
	>HSCRYABA	Human alpha-B-crystallin gene, 5' end.
	>HSCS5P	Human C3 gene, 5' end.

	>HSCSF1G1	Human gene for colony stimulating factor CSF-1 5' region
	>HSCSPA	Human cytotoxic serine proteinase gene, complete cds.
	>HSCST3G	Human CST3 gene for cystatin C
	>HSCST4	H.sapiens CST4 gene for Cystatin D
5	>HSCYP2C8	Human CYP2C8 gene for cytochrome P-450, 5' flank and exon 1
	>HSCYP45A	Human gene for cholesterol desmolase cytochrome P-450(SCC) exon 1
	>HSCYPB1	Human steroid 11-beta-hydroxylase (CYP11B1) gene, exons 1 and 2.
	>HSCYPXI	Human CYPXI gene for steroid 18-hydroxylase (P-450 C18). 2114
	>HSCYPXIB1	Human CYPXIB gene for steroid 11beta-hydroxylase (P-450 11beta),
10	>HSCYPXIX	Human CYPXIX gene, exon 1 coding for aromatase P-450 (EC 1.14.14.1)
	>HSDAFC1	Human decay-accelerating factor (DAF) gene, exons 1 and 2.
	>HSDBH1	Human DNA for dopamine beta-hydroxylase exon 1 (EC 1.14.17.1)
	>HSDDES	Human desmin gene, complete cds.
	>HSDKERB	Human cytokeratin 8 (CK8) gene, complete cds.
15	>HSDNAPOL	Human DNA polymerase alpha gene, 5' end.
	>HSDOPAM	H.sapiens dopamine D1A receptor gene, complete exon 1, and exon 2,
	>HSECP1	Human DNA for eosinophil cationic protein ECP
	>HSEGFA1	Human HER2 gene, promoter region and exon 1.
	>HSEL20	Human elastin gene, exon 1.
20	>HSELAM1B	Human endothelial leukocyte adhesion molecule I (ELAM-1) gene,

	>HSEMBPA	Human eosinophil major basic protein gene, complete cds.
	>HSENKB1	Human preproenkephalin B gene 5' region and exon 1
	>HSENO35	Human ENO3 gene 5' end for muscle--specific enolase
	>HSEOSDN	Human DNA for eosinophil derived neurotoxin
5	>HSEPR	Human erythropoietin receptor mRNA sequence derived from DNA, 5'
	>HSERB2P	Human c-erb B2/neu protein gene, 5'end, and promoter region.
	>HSERCC25	Human genomic and mRNA sequence for ERCC2 gene 5' region involved in
	>HSERPA	Human erythropoietin gene, complete cds.
	>HSERR	Human mRNA for oestrogen receptor
10	>HSESTE11	H.sapiens exon 1 for elastase I
	>HSFBRGG	Human gene for fibrinogen gamma chain
	>HSFCERG5	Human lymphocyte IgE receptor gene 5'-region (Fc-epsilon R)
	>HSFERG1	Human apoferritin H gene exon 1
	>HSFIBBR1	Human fibrinogen beta gene 5' region and exon 1
15	>HSFIXG	Human factor IX gene, complete cds.
	>HSFKBP1	Human FK506 binding proteins 12A, 12B and 12C (FKBP12) mRNA, exons
	>HSFLAP1	Human 5-lipoxygenase activating protein (FLAP) gene, exon 1.
	>HSFOS	Human fos proto-oncogene (c-fos), complete cds.
	>HSGOS2PE	Human GOS2 gene, upstream region and cds.
20	>HSGCSFG	Human gene for granulocyte colony--stimulating factor (G-CSF)
	>HSGEGR2	Human EGR2 gene 5' region 1233

	>HSGHPROM	Human growth hormone (hGH) gene promoter
	>HSGIPX1	Human gastric inhibitory polypeptide (GIP) mRNA, exon 1.
	>HSGLA	Human GLA gene for alpha-D-galactosidase A (EC 3.2.1.22)
	>HSGLUC1	Human glucagon gene transcription start region 732
5	>HSGMCSFG	Human gene for granulocyte-macrophage colony stimulating factor
	>HSGR1	Human glucocorticoid receptor gene, exon 1. 1602
	>HSGRFP1	Human growth hormone-releasing factor (GRF) gene, exon 1 (complete)
	>HSGSTP15	Human GST pi gene for glutathione S-transferase pi exon 1 to 5
	>HSGTRH	Human gene for gonadotropin-releasing hormone
10	>HSGYPC	Human glycophorin C (GPC) gene, exon 1, and promoter region.
	>HSH10	Human histone (H10) gene, 5' flank.
	>HSH1DNA	Human gene for H1 RNA 1057
	>HSH1FNC1	Human H1 histone gene FNC16 promoter region
	>HSH2B2H2	Human H2B.2 and H2A.1 genes for Histone H2A and H2B
15	>HSH4AHIS	H.sapiens H4/a gene for H4 histone
	>HSH4BHIS	H.sapiens H4/b gene for H4 histone
	>HSHARA	Human androgen receptor gene, transcription initiation sites.
	>HSHCG5B1	Human chorionic gonadotropin (hCG) beta subunit gene 5' 5'-flank
	>HSEMPRO	Human DNA for hemopoxin promoter
20	>HSHIAPPA	Human islet amyloid polypeptide (hIAPP) gene, complete cds.
	>HSHIH4	Human H4 histone gene

	>HSHISH2A	Human histone H2a gene
	>HSHISH2B	Human histone H2b gene
	>HSHISH3	Human histone H3 gene
	>HSHLAA1	Human HLA-A1 gene
5	>HSHLAB27	Human gene for HLA-B27 antigen
	>HSHLABW	Human HLA-Bw57 gene
	>HSHLAF	Human HLA-F gene for human leukocyte antigen F
	>HSHLIA	Human gene for histocompatibility antigen HLA-A3
	>HSHLIC	Human gene for class I histocompatibility antigen HLA-CW3
10	>HSHMG17G	Human HMG-17 gene for non-histone chromosomal protein HMG-17
	>HSHOX3D	Human HOX3D gene for homeoprotein HOX3D
	>HSHSC70	Human hsc70 gene for 71 kd heat shock cognate protein
	>HSHSP70D	Human heat shock protein (hsp 70) gene, complete cds.
	>HSHSP70P	Human hsp70B gene 5'-region
15	>HSIAPP12	Human IAPP gene exon 1 and exon 2 for islet amyloid polypeptide
	>HSICAMAB	Human intercellular adhesion molecule 1 (ICAM-1) gene, exon 1.
	>HSIFI54	Human interferon-inducible gene IFI-54K 5'flank
	>HSIFNA14	Human interferon alpha gene IFN-alpha 14
	>HSIFNA16	Human interferon alpha gene IFN-alpha 16
20	>HSIFNA5	Human interferon alpha gene IFN-alpha 5
	>HSIFNA6	Human interferon alpha gene IFN-alpha 6
	>HSIFNA7	Human interferon alpha gene IFN-alpha 7

	>HSIFNG	Human immune interferon (IFN-gamma) gene.
	>HSIFNIN6	Human al-pha/beta-interferon(IFN)-inducible 6-16 gene exon 1 and
	>HSIGF24B	Human DNA for insulin-like growth factor II (IGF-2); exon 4B
	>HSIGFBP1A	Human insulin-like growth factor binding protein (hIGFBP1) gene
5	>HSIGK10	Human germline gene for the leader peptide and variable region
	>HSIGK15	Human germline gene for the leader peptide and variable region
	>HSIGK17	Human rearranged gene for kappa immunoglobulin subgroup V kappa IV
	>HSIGK20	Human rearranged DNA for kappa immunoglobulin subgroup V kappa III
	>HSIGKLC1	Human germline fragment for immunoglobulin kappa light chain
10	>HSIGVA5	Human germline immunoglobulin kappa light chain V-segment
	>HSIL05	Human interleukin-2 (IL-2) gene and 5'-flanking region
	>HSIL1AG	Human gene for interleukin 1 alpha (IL-1 alpha)
	>HSIL1B	Human gene for prointerleukin 1 beta
	>HSIL2RG1	Human interleukin 2 receptor gene 5' flanking region and exon 1
15	>HSIL45	Human interleukin 4 gene 5'-region
	>HSIL5	Human interleukin 5 (IL-5) gene, complete cds.
	>HSIL6B	Human interleukin 6 (IL 6) gene, 5' flank.
	>HSIL71	Human interleukin 7 (IL7) gene, exon 1.
	>HSIL9A	Human IL9 protein gene, complete cds.

	>HSINSU	Human gene for preproinsulin, from chromosome 11. Includes a highly
	>HSINT1G	Human int-1 mammary oncogene
	>HSJUNCAA	Human jun-B gene, complete cds.
	>HSKER65A	Human DNA for 65 kD keratin type II exon 1 and 5' flank
5	>HSKERUHS	Human gene for ultra high-sulphur keratin protein
	>HSLACTG	Human alpha-lactalbumin gene
	>HSLAG1G	Human LAG-1 gene
	>HSLCATG	Human gene for lecithin-cholesterol acyltransferase (LCAT)
	>HSLCK1	Human lymphocyte-specific protein tyrosine kinase (lck) gene
10	>HSLFACD	Human leukocyte function-associated antigen-1 (LFA-1 or CD11a)
	>HSLPLA	Human lipoprotein lipase (LPL) gene, 5' flank.
	>HSLYAM01	Human leukocyte adhesion molecule-1 (LAM-1), exon 1.
	>HSLYSOZY	Human lysozyme gene (EC 3.2.1.17)
	>HSMBP1A	Human DNA for mannose binding protein 1 (MBP1), Exon 1
15	>HSMCCPAA	Human mast cell carboxypeptidase A (MC-CPA) gene, exons 1-2.
	>HSMDR1	Human P-glycoprotein (MDR1) mRNA, complete cds.
	>HSMED	Human bone marrow serine protease gene (medullasin)
	>HSMEHG	Human DNA (exon 1) for microsomal epoxide hydrolase
	>HSMETIE	Human metallothionein-Ie gene (hMT-Ie).
20	>HSMG01	Human myoglobin gene (exon 1)
	>HSMGSAG	Human gene for melanoma growth stimulatory activity (MGSA)

	>HSMHCAG1	Human alpha-MHC gene for myosin heavy chain N-terminus)
	>HSMHCGE1	Human class II invariant gamma-chain gene (5' flank, exon 1)
	>HSMHCW5	Human MHC class I HLA-Cw5 gene, 5' flank.
	>HSMLN1	Human motilin gene exon 1
5	>HSMPOA	Human myeloperoxidase gene, exons 1-4.
	>HSMRP	Human mitochondrial RNA-processing endoribonuclease RNA (mrp) gene
	>HSMTS1A	H.sapiens mts1 gene, 5' end.
	>HSMYCE12	Human myc-oncogene exon 1 and exon 2
	>HSNAKATP	Human Na,K-ATPase beta subunit (ATP1B) gene, exons 1 and 2.
10	>HSNEURK1	H.sapiens gene for neuromedin K receptor (exon 1)
	>HSNFH1	Human gene for heavy neurofilament subunit (NF-H) exon 1
	>HSNFIL6	Human gene for nuclear factor NF-IL6
	>HSNFLG	Human gene for neurofilament subunit NF-L
	>HSNK21	Human neurokinin-2 receptor (NK-2) gene, exon 1.
15	>HSNMYC	Human germ line n-myc gene
	>HSNRASPR	H. sapiens N-RAS promoter region
	>HSODC1A	Human ornithine decarboxylase (ODC1) gene, complete cds.
	>HSOTCEX1	Human ornithine transcarbamylase (OTC) gene, 5'-end region.
	>HSOTNPI	Human prepro-oxytocin-neurophysin I gene, complete cds.
20	>HSP450SCC	Human cytochrome P450scc gene, 5' end and promoter region.
	>HSP53G	Human p53 gene for transformation related protein p53

	>HSPADP	Human promoter DNA for Alzheimer's disease amyloid A4 precursor
	>HSPAI11	Human gene for plasminogen activator inhibitor 1 (PAI-1) 5'-flank
	>HSPGDF	Human platelet-derived growth factor A-chain (PDGF) gene, 5' end
	>HSPGP95G	Human PGP9.5 gene for neuron-specific ubiquitin C-terminal
5	>HSPLSM	Human plasminogen gene, exon 1.
	>HSPNMTB	Human gene for phenylethanolamine N-methylase (PNMT) (EC 2.1.1.28)
	>HSPOMC5F	Human opiomelanocortin gene, 5' flank.
	>HSPPI4B	Human placental protein 14 (PP14) gene, complete cds.
	>HSPRB3L	Human gene PRB3L for proline-rich protein G1
10	>HSPRB4S	Human PRB4 gene for proline-rich protein Po, allele S
	>HSPRLNC	Human prolactin mRNA, partial cds.
	>HSPROAA1	Human prothymosin-alpha gene, complete cds.
	>HSPROT2	Human protamine 2 gene, complete cds.
	>HSPRPE1	Human SPR2-1 gene for small proline rich protein (exon 1)
15	>HSPS2G1	Human estrogen-responsive gene pS2 5'flank and exon 1
	>HSPSAP	Human pulmonary surfactant apoprotein (PSAP) gene, complete cds.
	>HSPSP94A	Human gene for prostatic secretory protein PSP-94, exon 1
	>HSPTHRPA	Human parathyroid hormone-related peptide (PTHrP) gene, exons 1A,
	>HSPURNPHO	Human gene for purine nucleoside phosphorylase (upstream region)
20	>HSRDNA	Human rDNA origin of transcription

	>HSREGA01	Human regenerating protein (reg) gene, complete cds.
	>HSREN01	Human renin gene 5' region and exon 1
	>HSRPBG1	Human gene fragment for retinol binding protein (RBP) (exon 1-4)
	>HSSAA1A	Human serum amyloid A (GSAA1) gene, complete cds.
5	>HSSAA1B	H.sapiens SAA1 beta gene
	>HSSB4B1	Human gene fragment for HLA class II SB 4-beta chain (exon 1)
	>HSSISG5	Human c-sis proto-oncogene 5' region
	>HSSLIPG	Human SLPI gene for secretory leukocyte protease inhibitor
	>HSSOD1G1	Human superoxide dismutase (SOD-1) gene exon 1 and 5' flanking
10	>HSSODB	Human ornithine decarboxylase gene, complete cds.
	>HSSRDA01	H.sapiens steroid 5-alpha-reductase gene, exon 1.
	>HSSUBP1G	H.sapiens gene for substance P receptor (exon 1)
	>HSSYB1A1	Human synaptobrevin 1 (SYB1) gene, exon 1.
	>HSTAT1	Human gene for tyrosine aminotransferase (TAT) (EC 2.6.1.5) Exon 1.
15	>HSTCBV81	Human T-cell receptor V-beta 8.1 gene 775
	>HSTCRB21	Human T-cell receptor beta chain gene variable region.
	>HSTFG5	Human transferrin (Tf) gene 5' region
	>HSIL3FL5	Human interleukin 3 gene, 5' flank.
	>HSTFPB	Human tissue factor gene, complete cds.
20	>HSTGFB1	Human mRNA for transforming growth factor-beta (TGF-beta)

	>HSTGFB3B	Human transforming growth factor beta-3 gene, 5' end.
	>HSTGFBET2	Human transforming growth factor beta-2 gene, 5' end.
	>HSTH01	Human tyrosine hydroxylase (TH) (EC 1.14.16.2) gene from upstream
	>HSTHIO2A	Human metallothionein gene IIA promoter region
5	>HSTHRO01	Human thrombospondin gene, exons 1, 2 and 3.
	>HSTHXBG	H.sapiens gene for thyroxine-binding globulin gene
	>HSTHYR5	Human thyroglobulin gene 5' region
	>HSTNFA	Human gene for tumor necrosis factor (TNF-alpha)
	>HSTNFB	Human gene for lymphotoxin (TNF-beta)
10	>HSTOP01	Homo sapiens type I DNA topoisomerase gene, exons 1 and 2.
	>HSTPIA	Human triosephosphate isomerase (TPI) gene, 5' end.
	>HSTPO5	Human thyroid peroxidase gene 5' end (EC 1.11.1.7)
	>HSTRP	Human transferrin receptor gene promoter region
	>HSTRPY1B	Human tryptase-I gene, complete cds.
15	>HSTUBB2	Human beta 2 gene for beta-tubulin
	>HSTYRO1E	Human tyrosinase gene, exon 1 and 5' flanking region (EC 1.14.18.1)
	>HSU6RNA	Human gene for U 6 RNA
	>HSUPA	Human uPA gene for urokinase-plasminogen activator
	>HSVAVP01	Human proto-oncogene vav, 5' end.
20	>HSVCAM1A	Human vascular cell adhesion molecule-1 (VCAM1) gene, complete CDS.

>HSVIM5RR	Human vimentin gene 5' regulatory region
-----------	--

Once the gene target or, in the case of small pathogens, the genome target has been identified, short sequences within the gene or genome target are identified as medically significant target sites. Medically significant target sites can be defined as short DNA sequences (approximately 4-30 base pairs) that are required for the expression or replication of genetic material. For example, sequences that bind regulatory factors, either transcriptional or replicatory factors, are ideal target sites for altering gene or viral expression.

Further, coding sequences may be adequate target sites for disrupting gene function, although the disruption of a polymerase complex that is moving along the DNA sequence may require a stronger binder than for the disruption of the initial binding of a regulatory protein.

Finally, even non-coding, non-regulatory sequences may be of interest as target sites (e.g., for disrupting replication processes or introducing an increased mutational frequency).

Several specific examples of medically significant target sites are shown in Table IV.

Table IV
MEDICALLY SIGNIFICANT DNA-BINDING SEQUENCES

Test sequence	DNA-binding Protein	Medical Significance
EBV origin of replication	EBNA	infectious mononucleosis, nasal pharyngeal carcinoma
HSV origin of replication	UL9	oral and genital Herpes

Test sequence	DNA-binding Protein	Medical Significance
VZV origin of replication	UL9-like	shingles
HPV origin of replication	E2	genital warts, cervical carcinoma
Interleukin 2 enhancer	NFAT-1	immunosuppressant
HIV LTR	NFAT-1 NFkB	AIDS, ARC
HBV enhancer	HNF-1	hepatitis
Fibrogen promoter	HNF-1	cardiovascular disease
Oncogene promoter and coding sequences	??	cancer

15

(Abbreviations: EBV, Epstein-Barr virus; EBNA, Epstein-Barr virus nuclear antigen; HSV, Herpes Simplex virus; VZV, Varicella zoster virus; HPV, human papilloma virus; HIV LTR, Human immunodeficiency virus long terminal repeat; NFAT, nuclear factor of activated T cells; NFkB, nuclear factor kappaB; AIDS acquired immune deficiency syndrome; ARC, AIDS related complex; HBV, hepatitis B virus; HNF, hepatic nuclear factor.)

For example, origin of replication binding proteins have short, well-defined binding sites within the viral genome and are therefore excellent target sites for a competitive DNA-binding drug. Examples of such proteins include, Epstein Barr virus nuclear antigen 1 (EBNA-1) (Ambinder, et al.; Reisman, et al.), E2 (which is encoded by the human papilloma virus) (Chin, et al.), UL9 (which is encoded by herpes simplex virus type 1) (McGeoch, et al.), and the homologous protein in varicella zoster virus (VZV) (Stow, et al.).

Similarly, recognition sequences for DNA-binding proteins that act as transcriptional regulatory factors

are also good target sites for antiviral DNA-binding drugs. Examples listed in Table IV include (i) the binding site for hepatic nuclear factor (HNF-1), which is required for the expression of human hepatitis B virus (HBV) (Chang), and (ii) NF κ B and NFAT-1 binding sites in the human immunodeficiency virus (HIV) long terminal repeat (LTR), one or both of which may be involved in the expression of the virus (Greene, W.C.).

Examples of non-viral DNA targets for DNA-binding drugs are also shown in Table IV to illustrate the wide range of potential applications for sequence-specific DNA-binding molecules. For example, nuclear factor of activated T cells (NFAT-1) is a regulatory factor that is crucial to the inducible expression of the interleukin 2 gene in response to signals from the antigen receptor, which, in turn, is required for the cascade of molecular events during T cell activation (for review, see Edwards, C.A., and Crabtree, G.R.). The mechanism of action of two immunosuppressants, cyclosporin A and FK506, is thought to be to block the inducible expression of NFAT-1 (Schmidt, et al. and Banerji, et al.). However, the effects of these drugs are not specific to NFAT-1; therefore, a drug targeted specifically to the NFAT-1 binding site in the IL-2 enhancer would be desirable as an improved immunosuppressant.

Targeting the DNA site with a DNA-binding drug rather than targeting with a drug that affects the DNA-binding protein (presumably the target of the current immunosuppressants) is advantageous for at least two reasons: first, there are many fewer target sites for specific DNA sequences than specific proteins (e.g., in the case of glucocorticoid receptor, a handful of DNA-binding sites vs. about 50,000 protein molecules in each cell); and second, only the targeted gene need be

affected by a DNA-binding drug, while a protein-binding drug would disable all the cellular functions of the protein. An example of the latter point is the binding site for HNF-1 in the human fibrinogen promoter.

5 Fibrinogen level is one of the most highly correlated factor with cardiovascular disease. A drug targeted to either HNF-1 or the HNF-1 binding site in the fibrinogen promoter might be used to decrease fibrinogen expression in individuals at high risk for disease

10 because of the over-expression of fibrinogen. However, since HNF-1 is required for the expression of a number of normal hepatic genes, blocking the HNF-1 protein would be toxic to liver function. In contrast, by blocking a DNA sequence that is composed in part of the

15 HNF-1 binding site and in part by flanking sequences for divergence, the fibrinogen gene can be targeted with a high level of selectivity, without harm to normal cellular HNF-1 functions.

The assay has been designed to screen virtually

20 any DNA sequence. Test sequences of medical significance include viral or microbial pathogen genomic sequences and sequences within or regulating the expression of oncogenes or other inappropriately expressed cellular genes. In addition to the detection of

25 potential antiviral drugs, the assay of the present invention is also applicable to the detection of potential drugs for (i) disrupting the metabolism of other infectious agents, (ii) blocking or reducing the transcription of inappropriately expressed cellular

30 genes (such as oncogenes or genes associated with certain genetic disorders), and (iii) the enhancement or alteration of expression of certain cellular genes.

2. Defined Sets of Test Sequences.

The approach described in the above section emphasizes screening large numbers of fermentation broths, extracts, or other mixtures of unknowns against specific medically significant DNA target sequences. The assay can also be utilized to screen a large number of DNA sequences against known DNA-binding drugs to determine the relative affinity of the single drug for every possible defined specific sequence. For example, there are 4^n possible sequences, where n = the number of nucleotides in the sequence. Thus, there are $4^3 = 64$ different three base pair sequences, $4^4 = 256$ different four base pair sequences, $4^5 = 1024$ different 5 base pair sequences, etc. If these sequences are placed in the test site, the site adjacent to the screening sequence (the example used in this invention is the UL9 binding site), then each of the different test sequences can be screened against many different DNA-binding molecules.

The test sequences may be placed on either or both sides of the screening sequence, and the sequences flanking the other side of the test sequences are fixed sequences to stabilize the duplex and, on the 3' end of the top strand, to act as an annealing site for the primer (see Example 1). In Figure 14B, the TEST and SCREENING sequences are indicated. The preparation of such double-stranded oligonucleotides is described in Example 1 and illustrated in Figure 4.

The test sequences, denoted in Figure 14B as X:Y (where X = A,C,G, or T and Y = the complementary sequence, T,G,C, or A), may be any of the 256 different 4 base pair sequences shown in Figure 13.

Once a set of test oligonucleotides containing all possible four base pair sequences has been synthesized (see Example 1), the set can be screened with any DNA-

binding drug. The relative effect of the drug on each oligonucleotide assay system will reflect the relative affinity of the drug for the test sequence. The entire spectrum of affinities for each particular DNA sequence can therefore be defined for any particular DNA-binding drug. This data, generated using the assay of the present invention, can be used to facilitate molecular modeling programs and/or be used directly to design new DNA-binding molecules with increased affinity and specificity.

Another type of ordered set of oligonucleotides that may be useful for screening are sets comprised of scrambled sequences with fixed base composition. For example, if the recognition sequence for a protein is 5'-GATC-3' and libraries were to be screened for DNA-binding molecules that recognized this sequence, then it would be desirable to screen sequences of similar size and base composition as control sequences for the assay. The most precise experiment is one in which all possible 4 bp sequences are screened. In the case of a 4 base-pair sequence, this represents $4^4 = 256$ different test sequences: a number of screening sequences that may not be practical in every situation. However, there are many fewer possible 4 bp sequences with the same base composition (1G, 1A, 1T, 1C) ($n! = 24$ different 4 bp sequences with this particular base composition), such sequences provide excellent controls without having to screen large numbers of sequences.

30

3. Theoretical Considerations in Choosing Biological Target Sites: Specificity and Toxicity.

One consideration in choosing sequences to screen using the assay of the present invention is test

35

sequence accessibility, that is, the potential exposure of the sequence *in vivo* to binding molecules. Cellular DNA is packaged in chromatin, rendering most sequences relatively inaccessible. Sequences that are actively transcribed, particularly those sequences that are regulatory in nature, are less protected and more accessible to both proteins and small molecules. This observation is substantiated by a large literature on DNAase I sensitivity, footprinting studies with nucleases and small molecules, and general studies on chromatin structure (Tullius). The relative accessibility of a regulatory sequence, as determined by DNAase I hypersensitivity, is likely to be several orders of magnitude greater than an inactive portion of the cellular genome. For this reason the regulatory sequences of cellular genes, as well as viral regulatory or replication sequences, are useful regions to choose for selecting specific inhibitory small molecules using the assay of the present invention.

Another consideration in choosing sequences to be screened using the assay of the present invention is the uniqueness of the potential test sequence. As discussed above for the nuclear protein HNF-1, it is desirable that small inhibitory molecules are specific to their target with minimal cross reactivity. Both sequence composition and length effect sequence uniqueness. Further, certain sequences are found less frequently in the human genome than in the genomes of other organisms, for example, mammalian viruses. Because of base composition and codon utilization differences, viral sequences are distinctly different from mammalian sequences. As one example, the dinucleotide CG is found much less frequently in mammalian cells than the dinucleotide sequence GC: further, in SV40, a mammalian virus, the sequences AGCT and ACGT

are represented 34 and 0 times, respectively. Specific viral regulatory sequences can be chosen as test sequences keeping this bias in mind. Small inhibitory molecules identified which bind to such test sequences
5 will be less likely to interfere with cellular functions.

There are approximately 3×10^9 base pairs (bp) in the human genome. Of the known DNA-binding drugs for which there is crystallographic data, most bind 2-5 bp
10 sequences. There are $4^4 = 256$ different 4 base sequences; therefore, on average, a single 4 bp site is found roughly 1.2×10^7 times in the human genome. An individual 8 base site would be found, on average, about 50,000 times in the genome. On the surface, it
15 might appear that drugs targeted at even an 8 bp site might be deleterious to the cell because there are so many binding sites; however, several other considerations must be recognized.

First, most DNA is tightly wrapped in chromosomal
20 proteins and is relatively inaccessible to incoming DNA-binding molecules as demonstrated by the nonspecific endonucleolytic digestion of chromatin in the nucleus (Edwards, C.A. and Firtel, R.A.). Active transcription units are more accessible, but the most
25 highly exposed regions of DNA in chromatin are the sites that bind regulatory factors. As demonstrated by DNAase I hypersensitivity (Gross, D.S. and Garrard, W.T.), regulatory sites may be 100-1000 times more sensitive to endonucleolytic attack than the bulk of
30 chromatin. This is one reason for targeting regulatory sequences with DNA-binding drugs.

Secondly, several anticancer drugs that bind 2, 3, or 4 bp sequences have sufficiently low toxicity so that they can be used as drugs. This indicates that,
35 if high affinity binding sites for known drugs can be

matched with specific viral target sequences, it may be possible to use currently available drugs as antiviral agents at lower concentrations than they are currently used, with a concomitantly lower toxicity.

5

4. Further Considerations in Choosing Target Sites: Finding Eukaryotic Promoters.

Eukaryotic organisms have three RNA polymerases (Pol I, II, and III) that transcribe genetic information from DNA to RNA. The correct regulation of this information flow is essential for the survival of the cell. These multi-subunit enzymes need additional proteins to regulate transcription. Many of these additional proteins bind to DNA in a region 5' of the translation start site for a gene: this region is generally known as the promoter region of the gene.

All three polymerases use a core set of general transcription proteins to bind to this region. A central component of this complex is the protein called TBP or TFIID. The site this protein binds to is known as the TATA-box because the sequence usually contains a sequence motif similar to TATA (e.g., TATAa/tAa/t). Originally it was thought that each of the three polymerases used a separate set of general transcription factors and that Pol II used TFIID exclusively. Recently it has been shown that all three classes of RNA polymerase need TFIID for transcriptional regulation (see Comai, et al.; and Greenblatt)

A molecule that binds to a DNA sequence closely adjacent or overlapping a TATA binding site will likely alter transcriptional regulation of the gene. If the molecule binds based solely on specificity to the TATA-box sequence itself, then this molecule is expected to be very toxic to cells since the transcription of most

genes would be altered. The sequences adjacent to TATA boxes, however, are not conserved. Accordingly, if a particular sequence is selected adjacent a TATA box of a particular gene, a molecule that binds to this specific sequence would be expected to alter the transcriptional regulation of just that gene.

TATA-boxes were first identified by determining the sequence of the DNA located 5' of the RNA start sites of a number of genes. Examination of these sequences revealed that most genes had a TATA-box motif (consensus sequence) in the range of nucleotides 50 to 15 nucleotides 5' of the RNA start site. *In vitro* studies, typically DNA protection (footprinting) studies, lead to the conclusion that proteins were binding to these sites. Further *in vitro* DNA binding experiments demonstrated that some proteins could specifically bind to these sites. This lead to assays that allowed purification and subsequent sequencing of the binding proteins. This information facilitated the cloning and expression of genes encoding the binding proteins. A large number of transcription factors are now known. The protein designated TFIID has been demonstrated to bind to the TATA-box (Lee, et al.).

Molecules that interfere with the interaction of these transcription factors and their target DNA (*i.e.*, DNA/Protein transcription complexes) are also expected to alter transcription initiated from the target DNA. A publicly available database of these factors and the sequences to which they bind is available from the National Library of Medicine and is called "The Transcription Data Base, or TFD." The binding sites of these transcription factors can be identified in the 5' non-coding region of genes having known sequences (Example 15).

The ability to select target sequences adjacent the binding site of a transcription factor, as described above for TFIID, can be applied to other general transcription factors as well. For the purpose of the present invention, a general transcription factor is one that regulates the transcriptional expression of more than one gene. For any such general transcription factor, as for TFIID above, a particular target sequence can be selected adjacent the transcription factor binding site of a selected gene. A molecule that binds to this specific target sequence would be expected to alter the transcriptional regulation of just that gene and not all of the genes for which the transcription factor regulates expression. Alteration of transcriptional regulation may involve inhibition or increased affinity (enhancement) of binding of a transcription factor to its cognate DNA.

Many examples of such general transcription factors have been identified, including, but not limited to, the following: SP1 (Raney, et al., 1992; Kitadai, et al., 1992); NFAT-1 (Shaw, et al., 1988); Ets family of transcription factors, including Elf1 (Thompson, et al., 1992); Fos protein (Neuberg, et al., 1991); NF-kappa (Wirth, et al., 1988; Meijer, et al., 1992); and AP1-like proteins, including the product of the c-jun oncogene (Descheemaeker, et al., 1992; Ryder et al., 1988; Harshman et al., 1988; Angel et al., 1988; Bos et al., 1988; Bohmann et al., 1987).

Accordingly, for a selected gene, non-conserved DNA surrounding the transcription factor binding site can be chosen as a specific target sequence for small molecule binding. A small molecule can be chosen whose binding overlaps an adjacent transcription factor DNA binding sequence (e.g., by 1-3 nucleotide pairs). In this case, the specificity of DNA binding for the small

molecule is, in large part, derived from the non-conserved sequences adjacent the transcription factor binding site, in order to reduce small molecule binding at the transcription factor binding site associated with other genes.

Small molecules that bind such specific target sequences can be identified and/or designed using the assay and methods of the present invention.

10 5. Further Considerations in Choosing Alternative Small-Molecule-Binding Target Sites.

Small molecules that interfere with the interaction of any DNA binding protein and its cognate DNA (i.e., DNA/Protein complexes) can be selected by the assay and methods of the present invention. As described above for general transcription factors, sequences adjacent the DNA binding site for a selected DNA binding protein can serve as a target for small molecule binding in order to alter the interaction of the DNA binding protein and its cognate site. The small molecule can affect the DNA:protein interaction, for example, by inhibiting or enhancing the association of protein with the DNA.

25 For a selected DNA:protein interaction, non-conserved DNA surrounding the selected DNA binding site can be chosen as a specific target sequence for small molecule binding. In some cases the small molecule binding can overlap the DNA binding site: for example, in the case of a therapeutic used to treat a mammal with a bacterial infection, a small molecule may be selected to bind to the bacterial origin of DNA replication. Such a small molecule may essentially completely overlap the region defined by the bacterial origin-of-replication-DNA:protein interaction since a

corresponding target sequence is not likely present in the DNA of the mammalian host.

However, in the case where selective binding is required, as described above for TFIID, the specificity of the small molecule for DNA binding should essentially derive from the non-conserved sequences adjacent the DNA-binding protein's cognate DNA-binding site. This results in small molecule binding being reduced at similar DNA:protein binding sites at other locations.

10

6. Further Considerations in Choosing Target Sites: Procaryotes and Viruses.

Bacterial gene expression is regulated at several different levels, including transcription. General and specific transcription factors are needed along with the core RNA polymerase to accurately produce appropriate amounts of mRNA. Antibiotics that bind to the RNA polymerase and prevent mRNA production are potent bacterial poisons: molecules that could interfere with the initiation of transcription for specific essential genes are expected to have similar effects.

Many bacterial promoters have been sequenced and carefully examined. In general, the majority of bacterial promoters have two well characterized regions, the -35 region which has a consensus sequence similar to SEQ ID NO:625 and the -10 region with a consensus sequence of SEQ ID NO:626. The sequence of the start site for RNA polymerase, however, is not always the same. The start site is determined by a supplementary protein called the sigma factor, which confers specificity for binding the RNA polymerase core. Several sigma factors are present in any species of bacteria. Each sigma factor recognizes a different set of promoter sequences. Expression of sigma factors

is regulated, typically, by the growth conditions the bacteria is encountering. These sigma factor promoter sequences represent excellent targets for sequence specific DNA binding molecules.

5 As an example of choosing target sequences for the purpose of designing a DNA-binding therapeutic for a bacterial disease, consider the example of tuberculosis. Tuberculosis is caused by *Mycobacterium tuberculosis*.

10 All bacteria need to make ribosomes for the purpose of protein synthesis. The -35 and -10 regions of *M. tuberculosis* ribosome RNA synthesis has been determined. In the EMBL locus MTRRNOP the -35 signal is located at coordinants 394..400 and the -10 signal
15 is found at coordinants 419..422. These regions represent excellent targets for a DNA binding drug that would inhibit the growth of the bacteria by disrupting its ability to make ribosomes and synthesize protein. Multiple other essential genes could be targeted in a
20 similar manner.

M. tuberculosis is a serious public health problem for several reasons, including the development of antibiotic resistant strains. Many antibiotics inhibit the growth of bacteria by binding to a specific protein
25 and inhibiting its function. An example of this is the binding of rifampicin to the beta subunit of the bacterial RNA polymerase. Continued selection of bacteria with an agent of this kind can lead to the selection of mutants having an altered RNA polymerase
30 so that the antibiotic can no longer bind it. Such mutants can arise from a single mutation.

 However, binding a drug to a DNA regulatory region requires at least two mutations to escape the inhibitory effect of the drug: one mutation in the target DNA
35 sequence so that the drug could not bind the target se-

quence, and one mutation in the regulatory binding protein so that it can recognize the new, mutated regulatory sequence. Such a double mutation event is much less frequent than the single mutation discussed
5 above, for example, with rifampicin. Accordingly, it is expected that the development of drug resistant bacteria would be much less common for DNA-binding drugs that bind to promoter sequences.

The HIV viral promoter region (shown in Figure 28)
10 provides an example of choosing DNA target sequences for sequence-specific DNA binding drugs to inhibit viral replication.

Many eukaryotic viruses use promoter regions that have similar features to normal cellular genes. The
15 replication of these viruses depends on the general transcription factors present in the host cell. As such, the promoter sequences in DNA viruses are similar to those found in cellular genes and have been well-studied. The binding factors Sp-1 and TFIID are
20 important generalized factors that most viral promoters use.

In the HIV promoter sequence found in LOCUS HIVBH101 in version 32 of the EMBL databank, three tandem decanucleotide Sp1 binding sites are located
25 between positions 377 and 409. Site III shows the strongest affinity for the cellular factor. The three cause up to a tenfold effect on transcriptional efficiency in vitro. The transcription start site is at position 455, with a TATA box at 427-431 in the
30 sequence listed below. In addition to these sites, there are two NF-kappa-B sites in this region between nucleotides 350 and 373. These sites are annotated in Figure 28.

Sequence-specific DNA binding molecules that
35 specifically disrupted this binding would be expected

to disrupt HIV replication. For example, the sequences adjacent to the TFIID binding site (SEQ ID NO:628 and/or SEQ ID NO:629), would be target sites for a DNA-binding molecule designed to disrupt TFIID binding. These sequences are found in HIV but are not likely to occur overlapping TFIID binding sites in the endogenous human genome. Multiple sites could be targeted to decrease the likelihood that a single mutation could prevent drug binding.

D. Using Test Matrices and Pattern Matching for the Analysis of Data.

The assay described herein has been designed to use a single DNA:protein interaction to screen for sequence-specific and sequence-preferential DNA-binding molecules that can recognize virtually any specified sequence. By using sequences flanking the recognition site for a single DNA:protein interaction, a very large number of different sequences can be tested. The analysis of data yielded by such experiments displayed as matrices and analyzed by pattern matching techniques should yield information about the relatedness of DNA sequences.

The basic principle behind the DNA:protein assay of the present invention is that when molecules bind DNA sequences flanking the recognition sequence for a specific protein the binding of that protein is blocked. Interference with protein binding likely occurs by either (or both) of two mechanisms: (i) directly by steric hindrance, or (ii) indirectly by perturbations transmitted to the recognition sequence through the DNA molecule.

Both of these mechanisms will presumably exhibit distance effects. For inhibition by direct steric hindrance direct data for very small molecules is

available from methylation and ethylation interference studies. These data suggest that for methyl and ethyl moieties, the steric effect is limited by distance effects to 4-5 base pairs. Even still the number of
5 different sequences that can theoretically be tested for these very small molecules is still very large (i.e., 5 base pair combinations total 4^5 (=1024) different sequences).

In practice, the size of sequences tested can be
10 explored empirically for different sized test DNA-binding molecules. A wide array of sequences with increasing sequence complexity can be routinely investigated. This may be accomplished efficiently by synthesizing degenerate oligonucleotides and multiplex-
15 ing oligonucleotides in the assay process (i.e., using a group of different oligonucleotides in a single assay) or by employing pooled sequences in test matrices.

In view of the above, assays employing a specific
20 protein and oligonucleotides containing the specific recognition site for that protein flanked by different sequences on either side of the recognition site can be used to simultaneously screen for many different molecules, including small molecules, that have binding
25 preferences for individual sequences or families of related sequences. Figure 12 demonstrates how the analysis of a test matrix yields information about the nature of competitor sequence specificity. As an
30 example, to screen for molecules that could preferentially recognize each of the 256 possible tetranucleotide sequences (Figure 13), oligonucleotides could be constructed that contain these 256 sequences immediately adjacent to a 11 bp recognition sequence of UL9 *oriS* SEQ ID NO:615, which is identical in each construct.

145

In Figure 12 "+" indicates that the mixture retards or blocks the formation of DNA:protein complexes in solution and "-" indicates that the mixture had no marked effect on DNA:protein interactions. The results of this test are shown in Table V.

Table V

Test Mix	Specificity
#1,4,7: oligos	none detected for the above
#2: for recognition site	either nonspecific or specific
#3	AGCT
#5	CATT or ATT
#6	GCATTC, GCATT, CATTC, GCAT, or ATTC
#8	CTTT

These results demonstrate how such a matrix provides data on the presence of sequence specific binding activity in a test mixture and also provides inherent controls for non-specific binding. For example, the effect of test mix #8 on the different test assays reveals that the test mix preferentially affects the oligonucleotides that contain the sequence CCCT. Note that the sequence does not have to be within the test site for test mix #8 to exert an affect. By displaying the data in a matrix, the analysis of the sequences affected by the different test mixtures is facilitated.

Furthermore, defined, ordered sets of oligonucleotides can be screened with a chosen DNA-binding molecule. The results of these binding assays can then be examined using pattern matching techniques to determine the subsets of sequences that bind the molecule with similar binding characteristics. If the structural and

biophysical properties (such as, geometric shape and electrostatic properties) of sequences are similar, then it is likely that they will bind the molecule with similar binding characteristics. If the structural and biophysical properties of sequences are different, then it is likely that they will not bind the molecule with similar binding characteristics. In this context, the assay might be used to group defined, ordered sequences into subsets based on their binding characteristics: for example, the subsets could be defined as high affinity binding sites, moderate affinity binding sites, and low affinity binding sites. Sequences in the subsets with positive attributes (e.g., high affinity binding) have a high probability of having similar structural and biophysical properties to one another.

By screening and analyzing the binding characteristics of a number of DNA-binding molecules against the same defined set of DNA sequences, data can be accumulated about the subsets of sequences that fall into the same or similar subsets. Using this pattern matching approach, which can be computer-assisted, the sequences with similar structural and biophysical properties can be grouped empirically.

The database arising from pattern matching analysis of raw assay data will lead to the increased understanding of sequence structure and thereby lead to the design of novel DNA-binding molecules with related but different binding activities.

30

E. Applications for the Determination of the Sequence Specificity of DNA-Binding Drugs.

Applications for the determination of the sequence specificity of DNA-binding drugs are described below.

The applications are divided into drug homo- and

heteromeric polymers (part 1) and sequence-specific DNA-binding molecules as facilitators of triple strand formation (part 2).

One utility of the assay of the invention is the
5 identification of highest affinity binding sites among all possible sites of a certain length for a given DNA-binding molecule. This information may be valuable to the design of new DNA-binding therapeutics.

10 1. Multimerization of Sequence-Preferential or Sequence-Specific DNA-Binding Molecules Identified in the Assay.

Any particular DNA-binding small molecule screened in the assay may only recognize a 2-4 base
15 pair site, and even if the recognition is quite specific, the molecule may be toxic because there are so many target sites in the genome ($3 \times 10^9/4^4$ 4 bp sites, for example). However, if drugs with differential affinity for different sites are identified, the
20 toxicity of DNA-binding drugs may be drastically reduced by creating dimers, trimers, or multimers with these drugs (Example 13). From theoretical considerations of the free energy changes accompanying the binding of drugs to DNA, the intrinsic binding constant
25 of a dimer should be the square of the binding constant of the monomer (Le Pecq, J.B.). Experimental data confirmed this expectation in 1978 with dimer analogs of ethidium bromide (Kuhlmann, et al.). Dimerization of several intercalating molecules, in fact, yields
30 compounds with DNA affinities raised from 10^5 M⁻¹ for the corresponding monomer to 10^8 to 10^9 M⁻¹ for the dimers (Skorobogaty, et al.; Gaugain, et al. (1978a and b); Le Pecq, et al.; Pelaprat, et al.). Trimerization, which theoretically should yield binding
35 affinities that are the cube of the affinity of the

homomomeric subunit or the product of affinities of the heteromomeric subunits, has yielded compounds with affinities as high as 10^{12}M^{-1} (Laugaa, et al.). Such affinity is markedly better than the affinities
5 seen for many DNA regulatory proteins.

As a hypothetical example, if a relatively weak DNA-binding drug, drug X, which binds a 4 bp site with an affinity of $2 \times 10^5 \text{M}^{-1}$ was dimerized, the bis-X drug would now recognize an 8 bp site with a theoretical
10 affinity of $4 \times 10^{10} \text{M}^{-1}$. The difference in affinity between the monomer X and the bis-X form is 200,000-fold. The number of 4 bp sites in the genome is approximately 1.2×10^7 versus the number of 8 bp sites in the genome which is approximately 5×10^4 . Accord-
15 ingly, there are 256-fold fewer 8 bp sites than 4 bp sites. Thus, the number of high affinity target sites is 256-fold fewer for the bis-X molecule than the number of low affinity target sites for the monomer X, with a 200,000-fold difference in affinity between the
20 two types of sites.

Since the binding constant of a dimer is the product of the binding constants of the monomers, when monomers with higher initial binding constants are formed into dimers (or multimers) the differential
25 effect is proportionately increased, creating a wider "window" of affinity versus the number of binding sites. The breadth of the window essentially reflects the margin of effective drug concentration compared to the relative toxicity.

30 There are two immediate ramifications of dimerization (or multimerization) of monomeric drugs with moderate toxicity and sequence preference. First, the concentration of drug needed is lowered because of the higher affinity, so that even relatively toxic mole-
35 cules can be used as drugs. Second, since toxicity is

likely linked to the average number of drug molecules bound to the genome, as specificity is increased by increasing the length of the binding site, toxicity is decreased.

5 Given the information already available on sequence-preferential binding of DNA-binding drugs, it is likely that each drug presented to the screening assay will have (i) a number of high affinity binding sites (e.g., 10 to 100-fold better affinity than the average
10 site), (ii) a larger number of sites that are bound with moderate affinity (3 to 10-fold better affinity than average), (iii) the bulk of the binding sites having average affinity, and (iv) a number of sites having worse-than-average affinity. This range of
15 binding affinities will likely resemble a bell-shaped curve. The shape of the curve will probably vary for each drug. To exemplify, assume that approximately five 4 bp sites will be high affinity binding sites, and twenty 4 bp sites will be moderately high affinity
20 binding sites, then any given drug may recognize roughly 25, high or moderately high affinity binding sites. If 50 to 100 drugs are screened, this represents a "bank" of potentially 250-500 high affinity sites and 1000-2500 moderately high affinity sites.
25 Thus, the probability of finding a number of high affinity drug binding sites that match medically significant target sites is good. Furthermore, heterodimeric drugs can be designed to match DNA target sites of 8 or more bp, lending specificity to the
30 potential pharmaceuticals.

As discussed above, once the sequence preferences are known, the information may be used to design oligomeric molecules (homopolymers or heteropolymers) with substantially greater sequence specificity and
35 substantially higher binding affinity. For example, if

a DNA-binding molecule, X, binds a 4 bp sequence 5'-ACGT-3'/5'-ACGT-3' with an equilibrium affinity constant of $2 \times 10^5 \text{ M}^{-1}$, then the dimer of X, X_2 , should bind the dimer of the sequence, 5'-ACGTACGT-3'/5'-ACGTACGT-3', with an equilibrium affinity constant of $(2 \times 10^5 \text{ M}^{-1})^2 = 4 \times 10^{10} \text{ M}^{-2}$. The DNA-binding dimer molecule, X_2 , recognizes an 8 bp sequence, conferring higher sequence specificity, with a binding affinity that is theoretically 200,000-fold higher than the DNA-binding monomer, X.

The same argument can be extended to trimer molecules: the trimer of X, X_3 , would bind a 12 bp sequence, 5'-ACGTACGTACGT-3'/5'-ACGTACGTACGT-3', with a theoretical equilibrium affinity constant of $8 \times 10^{15} \text{ M}^{-2}$.

DNA-binding polymers constructed using the above-mentioned approach may be homo- or hetero-polymers of the parent compounds or oligomeric compounds composed of mixed subunits of the parent compounds. Homopolymers are molecules constructed using two or more subunits of the same monomeric DNA-binding molecule. Heteropolymers are molecules constructed using two or more subunits of different monomeric DNA-binding molecules. Oligomeric compounds are constructed of mixed pieces of parent compounds and may be hetero- or homomeric.

For example, distamycin is a member of a family of non-intercalating minor groove DNA-binding oligopeptides that are composed of repeating units of N-methylpyrrole groups. Distamycin has 3 N-methylpyrrole groups. Examples of homopolymers would be bis-distamycin, the dimer of distamycin, a molecule containing 6 N-methylpyrrole groups or tris-distamycin, the trimer of distamycin, a molecule containing 9 N-methylpyrrole groups.

Daunomycin is a member of an entirely different class of DNA-binding molecules, the anthracycline antibiotics, that bind to DNA via intercalation. Heteropolymers are molecules composed of different types of DNA-binding subunits; for example, compounds composed of a distamycin molecule linked to a daunomycin molecule or a distamycin molecule linked to two daunomycin molecules. The term "oligomeric" is being used to describe molecules comprised of linked subunits each of which may be smaller than the parent compound.

An example of an homo-oligomeric compound would be a distamycin molecule linked to 1 or 2 additional N-methylpyrrole groups; the resulting molecule would not be as large as bis-distamycin, but would fundamentally be composed of the same component organic moieties that comprise the parent molecule. Examples of a hetero-oligomeric compounds would be daunomycin linked to one or two N-methylpyrrole groups.

The construction of these polymers will be directed by the information derived from the sequence preferences of the parent compounds tested in the assay. In one embodiment of the assay, a database of preferred sequences is constructed, providing a source of information about the 4 bp sequences that bind with relatively higher affinity to particular drugs that may be linked together to target any particular larger DNA sequence.

DNA-binding subunits can be chemically coupled to form heteropolymers or homopolymers. The subunits can be joined directly to each other, as in the family of distamycin molecules, or the subunits can be joined with a spacer molecule, such as carbon chains or peptide bonds. The coupling of subunits is dependent on the chemical nature of the subunits: appropriate coupling reactions can be determined for any two

subunit molecules from the chemical literature. The choice of subunits will be directed by the sequence to be targeted and the data accumulated through the methods discussed in Section VI.B of this application.

5

2. Sequence-Specific DNA-Binding Molecules Identified in the Assay as Facilitators of Triplex Formation.

Several types of nucleic acid base-containing polymers have been described that will form complexes with nucleic acids (for reviews, see Helene, C. and Toulme, J.-J.). One type of such a polymer forms a triple-stranded complex by the insertion of a third strand into the major groove of the DNA helix. Several types of base-recognition specific interactions of third strand oligonucleotide-type polymers have been observed. One type of specificity is due to Hoogsteen bonding (Hoogsteen). This specificity arises from recognition between pyrimidine oligonucleotides and double-stranded DNA by pairing thymine and adenine:thymine base pairs and protonated cytosine and guanine:cytosine base pairs (Griffin, et al.). Another type of specific interaction involves the use of purine oligonucleotides for triplex formation. In these triplexes, adenine pairs with adenine:thymine base pairs and guanine with guanine:cytosine (Cooney, et al.; Beal and Dervan) or thymine:adenine base pairs (Griffen, L., and Dervan, P.B.).

Other motifs for triplex formation have been described, including the incorporation of nucleic acid analogs (eg, methylphosphonates, phosphorothioates; Miller, et al.), and the invention of backbones other than the phosphoribose backbones normally found in nucleic acids (Pitha, et al.; Summerton, et al.). In several cases, the formation of triplex has been

demonstrated to inhibit the binding of a DNA-binding protein (e.g., Young, et al.; Maher, et al.) or the expression of a cellular protein (Cooney, et al.).

Furthermore, several experiments have been reported in which a small DNA-binding molecule has been covalently attached to polymer capable of forming a triplex structure: (i) an acridine:polypyrimidine molecule has been demonstrated to inhibit SV40 in CV-1 cells (Birg, et al.); (ii) cleavage at a single site in a yeast chromosome was achieved with an oligonucleotide:EDTA-Fe molecule (Strobel, et al.; Dervan); and (iii) a photoinducible endonuclease was created by similar strategy by attaching an ellipticine derivative to a homopyrimidine oligonucleotide (Perouault, et al.). Several other small intercalating agents coupled to oligonucleotides have been described (for review, see Montenay-Garestier).

One utility of the assay of the present invention is to identify the sequence-specificity of DNA-binding molecules for use in designing and synthesizing heteromeric therapeutics consisting of a DNA-binding polymer (e.g., an oligonucleotide) attached to a sequence-preferential or sequence-specific DNA-binding molecule, yielding a heteropolymer. The attached small molecule may serve several functions.

First, if the molecule has increased affinity for a specific site (such as, a particular 4 base pair sequence) over all other sites of the same size, then the local concentration of the hetero-molecule, including the oligonucleotide, will be increased at those sites. The amount of heteropolymer, containing a sequence-specific moiety attached to one end, needed for treatment purposes is reduced compared to a heteropolymer that has a non-specific DNA-binding moiety attached. This reduction in treatment amount is directly

proportional to both the differential specificity and the relative affinities between the sequence-specific binder and the non-specific binder. For the simplest example, if a sequence-specific molecule with absolute
5 specificity (i.e., it binds only one sequence) had equal affinity for a specific 4 base-pair target site (1/256 possible combinations) as a non-specific molecule, then the amount of drug needed to exert the same effective concentration at that site could potentially
10 be as much as 256-fold less for the specific and non-specific drugs. Accordingly, attaching a sequence-specific DNA-binding molecule to a polymer designed to form triplex structures allows increased localized concentrations.

15 A second utility of the assay of the present invention is to identify small molecules that cause conformational changes in the DNA when they bind. The formation of triplex DNA requires a shift from B form to A form DNA. This is not energetically favorable,
20 necessitating the use of increased amounts of polymer for triplex formation to drive the conformational change. However, the insertion of a small DNA-binding molecule (such as, actinomycin D), which induces a conformational change in the DNA, reduces the amount of
25 polymer needed to stabilize triplex formation.

Accordingly, one embodiment of the invention is to use the assay to test known DNA-binding molecules with all 256 possible four base pair test sequences to determine the relative binding affinity to all possible
30 4 bp sequences. Then, once the sequence preferences are known, the information may be used to design hetercopolymeric molecules comprised of a small DNA-binding molecule and a macromolecule, such as a triplex-forming oligonucleotide, to obtain a DNA-
35 binding molecule with enhanced binding characteristics.

The potential advantages of attaching a sequence-specific or sequence-preferential DNA-binding small molecule to a triplex forming molecule are to (i) target the triplex to a subset of specific DNA sequences and thereby (ii) anchor the triplex molecule in the vicinity of its target sequence and in doing so, (iii) increase the localized concentration of the triplex molecule, which allows (iv) lower concentrations of triplex to be used effectively. The presence of the small molecule may also facilitate localized perturbations in DNA structure, such as destabilizing the B form of DNA, which is unsuitable for triplex formation. Such destabilization may facilitate the formation of other structures, such a form DNA useful for triplex formation. The net effect would be to decrease the amount of triplex needed for efficacious results.

F. Other Applications.

The potential pharmaceutical applications for sequence-specific DNA-binding molecules are very broad, including antiviral, antifungal, antibacterial, antitumor agents, immunosuppressants, and cardiovascular drugs. Sequence-specific DNA-binding molecules can also be useful as molecular reagents as, for example, specific sequence probes.

As more DNA-binding molecules are detected, information about their DNA binding affinities, sequence recognition, and mechanisms of DNA-binding will be gathered, eventually facilitating the design and/or modification of new molecules with different or specialized activities.

Although the assay has been described in terms of the detection of sequence-specific DNA-binding molecules, the reverse assay could be achieved by adding

156

DNA in excess to protein to look for peptide sequence specific protein-binding inhibitors.

The following examples illustrate, but in no way are intended to limit, the present invention.

Materials and Methods

Synthetic oligonucleotides were prepared using commercially available automated oligonucleotide synthesizers. Alternatively, custom designed synthetic oligonucleotides may be purchased, for example, from Synthetic Genetics (San Diego, CA). Complementary strands were annealed to generate double-strand oligonucleotides.

Restriction enzymes were obtained from Boehringer Mannheim (Indianapolis IN) or New England Biolabs (Beverly MA) and were used as per the manufacturer's directions.

Distamycin A and Doxorubicin were obtained from Sigma (St. Louis, MO). Actinomycin D was obtained from Boehringer Mannheim or Sigma.

Standard cloning and molecular biology techniques are described in Ausubel, et al., and Sambrook, et al.

25

Example 1

Preparation of the Oligonucleotide Containing the Screening Sequence

This example describes the preparation of (A) biotinylated/digoxigenin/radiolabeled, and (B) radio-labeled double-stranded oligonucleotides that contain the screening sequence and selected Test sequences.

A. Biotinylation.

The oligonucleotides were prepared as described above. The wild-type control sequence for the UL9

binding site, as obtained from HSV, is shown in Figure 4. The screening sequence, *i.e.* the UL9 binding sequence, is CGTTCGCACTT (SEQ ID NO:601) and is underlined in Figure 4. Typically, sequences 5' and/or 3' to the screening sequence were replaced by a selected Test sequence (Figure 5).

One example of the preparation of a site-specifically biotinylated oligonucleotide is outlined in Figure 4. An oligonucleotide primer complementary to the 3' sequences of the screening sequence-containing oligonucleotide was synthesized. This oligonucleotide terminated at the residue corresponding to the C in position 9 of the screening sequence. The primer oligonucleotide was hybridized to the oligonucleotide containing the screening sequence. Biotin-11-dUTP (Bethesda Research Laboratories (BRL), Gaithersburg MD) and Klenow enzyme were added to this complex (Figure 4) and the resulting partially double-stranded biotinylated complexes were separated from the unincorporated nucleotides using either pre-prepared "G-25 SEPHADEX" spin columns (Pharmacia, Piscataway NJ) or "NENSORB" columns (New England Nuclear) as per manufacturer's instructions. The remaining single-strand region was converted to double-strands using DNA polymerase I Klenow fragment and dNTPs resulting in a fully double-stranded oligonucleotide. A second "G-25 SEPHADEX" column was used to purify the double-stranded oligonucleotide. Oligonucleotides were diluted or resuspended in 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 mM EDTA and stored at -20°C. For radiolabelling the complexes, ³²P-alpha-dCTP (New England Nuclear, Wilmington, DE) replaced dCTP for the double-strand completion step.

Alternatively, the top strand, the primer, or the fully double-stranded oligonucleotide have been radiolabeled with γ -³²P-ATP and polynucleotide kinase

(NEB, Beverly, MA). Most of our preliminary studies have employed radiolabeled, double-stranded oligonucleotides. The oligonucleotides are prepared by radiolabeling the primer with T4 polynucleotide kinase and γ -³²P-ATP, annealing the "top" strand full length oligonucleotide, and "filling-in" with Klenow fragment and deoxynucleotide triphosphates. After phosphorylation and second strand synthesis, oligonucleotides are separated from buffer and unincorporated triphosphates using "G-25 SEPHADEX" preformed spin columns (IBI, New Haven, CT or Biorad, Richmond CA). This process is outlined in Figure 4. The reaction conditions for all of the above Klenow reactions were as follows: 10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 50 mM NaCl, 1 mM dithioerythritol, 0.33-100 μ M deoxytriphosphates, 2 units Klenow enzyme (Boehringer-Mannheim, Indianapolis IN). The Klenow reactions were incubated at 25°C for 15 minutes to 1 hour. The polynucleotide kinase reactions were incubated at 37°C for 30 minutes to 1 hour.

B. End-Labeling with Digoxigenin.

The biotinylated, radiolabelled oligonucleotides or radiolabeled oligonucleotides were isolated as above and resuspended in 0.2 M potassium cacodylate (pH=7.2), 4 mM MgCl₂, 1 mM 2-mercaptoethanol, and 0.5 mg/ml bovine serum albumin. To this reaction mixture digoxigenin-11-dUTP (an analog of dTTP, 2'-deoxyuridine-5'-triphosphate, coupled to digoxigenin via an 11-atom spacer arm, Boehringer Mannheim, Indianapolis IN) and terminal deoxynucleotidyl transferase (GIBCO BRL, Gaithersburg, MD) were added. The number of Dig-11-dUTP moieties incorporated using this method appeared to be less than 5 (probably only 1 or 2) as judged by electrophoretic mobility on polyacrylamide

gels of the treated fragment as compared to oligonucleotides of known length.

The biotinylated or non-biotinylated, digoxigenin-containing, radiolabelled oligonucleotides were isolated as above and resuspended in 10 mM Tris-HCl, 1 mM EDTA, 50 mM NaCl, pH 7.5 for use in the binding assays.

The above procedure can also be used to biotinylate the other strand by using an oligonucleotide containing the screening sequence complementary to the one shown in Figure 4 and a primer complementary to the 3' end of that molecule. To accomplish the biotinylation Biotin-7-dATP was substituted for Biotin-11-dUTP. Biotinylation was also accomplished by chemical synthetic methods: for example, an activated nucleotide is incorporated into the oligonucleotide and the active group is subsequently reacted with NHS-LC-Biotin (Pierce). Other biotin derivatives can also be used.

20 C. Radiolabelling the Oligonucleotides.

Generally, oligonucleotides were radiolabelled with gamma-³²P-ATP or alpha-³²P-deoxynucleotide triphosphates and T4 polynucleotide kinase or the Klenow fragment of DNA polymerase, respectively. Labelling reactions were performed in the buffers and by the methods recommended by the manufacturers (New England Biolabs, Beverly MA; Bethesda Research Laboratories, Gaithersburg MD; or Boehringer/Mannheim, Indianapolis IN). Oligonucleotides were separated from buffer and unincorporated triphosphates using "G-25 SEPHADEX" preformed spin columns (IBI, New Haven, CT; or Biorad, Richmond, CA) or "NENSORB" preformed columns (New England Nuclear, Wilmington, DE) as per the manufacturers instructions.

160

There are several reasons to enzymatically synthesize the second strand. The two main reasons are that by using an excess of primer, second strand synthesis can be driven to near completion so that nearly all top strands are annealed to bottom strands, which prevents the top strand single strands from folding back and creating additional and unrelated double-stranded structures, and secondly, since all of the oligonucleotides are primed with a common primer, the primer can bear the end-label so that all of the oligonucleotides will be labeled to exactly the same specific activity.

Example 2

15

Preparation of the UL9 Protein

A. Cloning of the UL9 Protein-Coding Sequences into pAC373.

To express full length UL9 protein a baculovirus expression system has been used. The sequence of the UL9 coding region of Herpes Simplex Virus has been disclosed by McGeoch et al. and is available as an EMBL nucleic acid sequence. The recombinant baculovirus AcNPV/UL9A, which contained the UL9 protein-coding sequence, was obtained from Mark Challberg (National Institutes of Health, Bethesda MD). The construction of this vector has been previously described (Olivo, et al. (1988, 1989)). Briefly, the *NarI*/*EcoRV* fragment was derived from pMC160 (Wu, et al.). Blunt-ends were generated on this fragment by using all four dNTPs and the Klenow fragment of DNA polymerase I (Boehringer Mannheim, Indianapolis IN) to fill in the terminal overhangs. The resulting fragment was blunt-end ligated into the unique *BamHI* site of the baculoviral vector pAC3T3 (Summers, et al.).

35

B. Cloning of the UL9 Sequence in pVL1393.

The UL9 protein-coding region was cloned into a second baculovirus vector, pVL1393 (Luckow, et al.). The 3077 bp *NarI/EcoRV* fragment containing the UL9 gene was excised from vector pEcoD (obtained from Dr. Bing Lan Rong, Eye Research Institute, Boston, MA): the plasmid pEcoD contains a 16.2 kb *EcoRI* fragment derived from HSV-I that bears the UL9 gene (Goldin, et al.). Blunt-ends were generated on the UL9-containing fragment as described above. *EcoRI* linkers (10 mer) were blunt-end ligated (Ausubel, et al.; Sambrook, et al.) to the blunt-ended *NarI/EcoRV* fragment.

The vector pVL1393 (Luckow, et al.) was digested with *EcoRI* and the linearized vector isolated. This vector contains 35 nucleotides of the 5' end of the coding region of the polyhedron gene upstream of the polylinker cloning site. The polyhedron gene ATG has been mutated to ATT to prevent translational initiation in recombinant clones that do not contain a coding sequence with a functional ATG. The *EcoRI/UL9* fragment was ligated into the linearized vector, the ligation mixture transformed into *E. coli* and ampicillin resistant clones selected. Plasmids recovered from the clones were analyzed by restriction digestion and plasmids carrying the insert with the amino terminal UL9 protein-coding sequences oriented to the 5' end of the polyhedron gene were selected. This plasmid was designated pVL1393/UL9 (Figure 7).

pVL1393/UL9 was cotransfected with wild-type baculoviral DNA (AcMNPV; Summers, et al.) into Sf9 (*Spodoptera frugiperda*) cells (Summers, et al.). Recombinant baculovirus-infected Sf9 cells were identified and clonally purified (Summers, et al.).

C. Expression of the UL9 Protein.

Clonal isolates of recombinant baculovirus infected Sf9 cells were grown in Grace's medium as described by Summers, et al. The cells were scraped
5 from tissue culture plates and collected by centrifugation (2,000 rpm, for 5 minutes, 4°C). The cells were then washed once with phosphate buffered saline (PBS) (Maniatis, et al.). Cell pellets were frozen at -70°C. For lysis the cells were resuspended in 1.5 volumes 20
10 mM HEPES, pH 7.5, 10% glycerol, 1.7 M NaCl, 0.5 mM EDTA, 1 mM dithiothreitol (DTT), and 0.5 mM phenyl methyl sulfonyl fluoride (PMSF). Cell lysates were cleared by ultracentrifugation (Beckman table top ultracentrifuge, TLS 55 rotor, 34 krpm, 1 hr, 4°C).
15 The supernatant was dialyzed overnight at 4°C against 2 liters dialysis buffer (20 mM HEPES, pH 7.5, 10% glycerol, 50 mM NaCl, 0.5 mM EDTA, 1 mM dtt, and 0.1 mM PMSF).

These partially purified extracts were prepared
20 and used in DNA:protein binding experiments. If necessary extracts were concentrated using a "CENTRICON 30" filtration device (Amicon, Danvers MA).

D. Cloning the Truncated UL9 Protein.

25 The sequence encoding the C-terminal third of UL9 and the 3' flanking sequences, an approximately 1.2 kb fragment, was subcloned into the bacterial expression vector, pGEX-2T (Figure 6). The pGEX-2T is a modification of the pGEX-1 vector of Smith, et al. which
30 involved the insertion of a thrombin cleavage sequence in-frame with the glutathione-S-transferase protein (gst).

A 1,194 bp *Bam*HI/*Eco*RV fragment of pEcoD was isolated that contained a 951 bp region encoding the C-

terminal 317 amino acids of UL9 and 243 bp of the 3' untranslated region.

This *Bam*HI/*Eco*RV UL9 carboxy-terminal (UL9-COOH) containing fragment was blunt-ended and *Eco*RI linkers added as described above. The *Eco*RI linkers were designed to allow in-frame fusion of the UL9 protein-coding sequence to the *gst*-thrombin coding sequences. The linkered fragment was isolated and digested with *Eco*RI. The pGEX-2T vector was digested with *Eco*RI, treated with Calf Intestinal Alkaline Phosphatase (CIP) and the linear vector isolated. The *Eco*RI linkered UL9-COOH fragment was ligated to the linear vector (Figure 6). The ligation mixture was transformed into *E. coli* and ampicillin resistant colonies were selected. Plasmids were isolated from the ampicillin resistant colonies and analyzed by restriction enzyme digestion. A plasmid which generated a *gst*/thrombin/-UL9-COOH in frame fusion was identified (Figure 6) and designated pGEX-2T/UL9-COOH.

20

E. Expression of the Truncated UL9 Protein.

E. coli strain JM109 was transformed with pGEX-2T/C-UL9-COOH and was grown at 37°C to saturation density overnight. The overnight culture was diluted 1:10 with LB medium containing ampicillin and grown from one hour at 30°C. IPTG (isopropylthio- β -galactoside) (GIBCO-BRL) was added to a final concentration of 0.1 mM and the incubation was continued for 2-5 hours. Bacterial cells containing the plasmid were subjected to the temperature shift and IPTG conditions, which induced transcription from the *tac* promoter.

30

Cells were harvested by centrifugation and resuspended in 1/100 culture volume of MTPBS (150 mM NaCl, 16 mM Na₂HPO₄, 4 mM NaH₂PO₄). Cells were lysed by

sonication and lysates cleared of cellular debris by centrifugation.

The fusion protein was purified over a glutathione agarose affinity column as described in detail by Smith, et al. The fusion protein was eluted from the affinity column with reduced glutathione, dialyzed against UL9 dialysis buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.1 mM PMSF) and cleaved with thrombin (2 ng/ug of fusion protein).

An aliquot of the supernatant obtained from IPTG-induced cultures of pGEX-2T/C-UL9-COOH-containing cells and an aliquot of the affinity-purified, thrombin-cleaved protein were analyzed by SDS-polyacrylamide gel electrophoresis. The result of this analysis is shown in Figure 8. The 63 kilodalton GST/C-UL9 fusion protein is the largest band in the lane marked GST-UL9 (lane 2). The first lane contains protein size standards. The UL9-COOH protein band (lane GST-UL9 + Thrombin, Figure 8, lane 3) is the band located between 30 and 46 kD: the glutathione transferase protein is located just below the 30 kD size standard. In a separate experiment a similar analysis was performed using the uninduced culture: it showed no protein corresponding in size to the fusion protein.

Extracts are dialyzed before use. Also, if necessary, the extracts can be concentrated typically by filtration using a "CENTRICON 30" filter.

Example 3

Binding Assays

A. Band Shift Gels.

DNA:protein binding reactions containing both labelled complexes and free DNA were separated electrophoretically on 4-10% polyacrylamide/Tris-Borate-EDTA (TBE) gels (Fried, et al.; Garner, et al.). The gels

were then fixed, dried, and exposed to X-ray film. The autoradiograms of the gels were examined for band shift patterns.

5 B. Filter Binding Assays.

A second method used particularly in determining the half-lives for oligonucleotide:protein complexes is filter binding (Woodbury, et al.). Nitrocellulose disks (Schleicher and Schuell, BA85 filters) that have
10 been soaked in binding buffer (see below) were placed on a vacuum filter apparatus. DNA:protein binding reactions (see below; typically 15-30 μ l) are diluted to 0.5 ml with binding buffer (this dilutes the concentration of components without dissociating
15 complexes) and applied to the discs with vacuum applied. Under low salt conditions the DNA:protein complex sticks to the filter while free DNA passes through. The discs are placed in scintillation counting fluid (New England Nuclear), and the cpm
20 determined using a scintillation counter.

This technique has been adapted to 96-well and 72-slot nitrocellulose filtration plates (Schleicher and Schuell) using the above protocol except (i) the reaction dilution and wash volumes are reduced, and
25 (ii) the flow rate through the filter is controlled by adjusting the vacuum pressure. This method greatly facilitates the number of assay samples that can be analyzed. Using radioactive oligonucleotides, the samples are applied to nitrocellulose filters, the
30 filters are exposed to x-ray film, then analyzed using a Molecular Dynamics scanning densitometer. This system transfers data directly into analytical software programs (e.g., Excel) for analysis and graphic display.

Example 4Functional UL9 Binding AssayA. Functional DNA-Binding Activity Assay.

Purified protein was tested for functional activity using band-shift assays. Radiolabelled oligonucleotides (prepared as in Example 1B) that contain the 11 bp recognition sequence were mixed with the UL9 protein in binding buffer (optimized reaction conditions: 0.1 ng ³²P-DNA, 1 ul UL9 extract, 20 mM HEPES, pH 7.2, 50 mM KCl, and 1 mM DTT). The reactions were incubated at room temperature for 10 minutes (binding occurs in less than 2 minutes), then separated electrophoretically on 4-10% non-denaturing polyacrylamide gels. UL9-specific binding to the oligonucleotide is indicated by a shift in mobility of the oligonucleotide on the gel in the presence of the UL9 protein but not in its absence. Bacterial extracts containing (+) or without (-) UL9 protein and affinity purified UL9 protein were tested in the assay. Only bacterial extracts containing UL9 or affinity purified UL9 protein generate the gel band-shift indicating protein binding.

The degree of extract that needed to be added to the reaction mix, in order to obtain UL9 protein excess relative to the oligonucleotide, was empirically determined for each protein preparation/extract. Aliquots of the preparation were added to the reaction mix and treated as above. The quantity of extract at which the majority of the labelled oligonucleotide appears in the DNA:protein complex was evaluated by band-shift or filter binding assays. The assay is most sensitive under conditions in which the minimum amount of protein is added to bind most of the DNA. Excess protein decreases the sensitivity of the assay with respect to the ability of inhibitors to compete with

the protein for oligonucleotide binding, except when protein concentrations are so high that non-specific protein/DNA binding is provoked.

5 B. Rate of Dissociation.

The rate of dissociation is determined using a competition assay. An oligonucleotide having the sequence presented in Figure 4, which contained the binding site for UL9 (SEQ ID NO:614), was radiolabelled
10 with ^{32}P -ATP and polynucleotide kinase (Bethesda Research Laboratories). The competitor DNA was a 17 base pair oligonucleotide (SEQ ID NO:616) containing the binding site for UL9.

In the competition assays, the binding reactions
15 (Example 4A) were assembled with each of the oligonucleotides and placed on ice. Unlabelled oligonucleotide (1 μg) was added 1, 2, 4, 6, or 21 hours before loading the reaction on an 8% polyacrylamide gel (run in TBE buffer (Maniatis, et al.)) to separate the
20 reaction components. The dissociation rates, under these conditions, for the truncated UL9 (UL9-COOH) and the full length UL9 is approximately 4 hours at 4°C. In addition, random oligonucleotides (a 10,000-fold excess) that did not contain the UL9 binding sequence
25 and sheared herring sperm DNA (a 100,000-fold excess) were tested: neither of these control DNAs competed for binding with the oligonucleotide containing the UL9 binding site.

30 C. Optimization of the UL9 Binding Assay.

1. Truncated UL9 from the Bacterial Expression System.

The effects of the following components on the binding and dissociation rates of UL9-COOH with its
35 cognate binding site have been tested and optimized:

buffering conditions (including the pH, type of buffer, and concentration of buffer); the type and concentration of monovalent cation; the presence of divalent cations and heavy metals; temperature; various
5 polyvalent cations at different concentrations; and different redox reagents at different concentrations. The effect of a given component was evaluated starting with the reaction conditions given above and based on the dissociation reactions described in Example 4B.

10 The optimized conditions used for the binding of UL9-COOH contained in bacterial extracts (Example 2E) to oligonucleotides containing the HSV ori sequence (SEQ ID NO:601) were as follows: 20 mM HEPES, pH 7.2, 50 mM KCl, 1 mM DTT, 0.005 - 0.1 ng radiolabeled
15 (specific activity, approximately 10^8 cpm/ μ g) or digoxigenated, biotinylated oligonucleotide probe, and 5-10 μ g crude UL9-COOH protein preparation (1 mM EDTA is optional in the reaction mix). Under optimized conditions, UL9-COOH binds very rapidly and has a
20 dissociation rate of about 4 hours at 4°C with non-biotinylated oligonucleotide and 5-10 minutes with biotinylated oligonucleotides. The dissociation rate of UL9-COOH changes markedly under different physical conditions. Typically, the activity of a UL9 protein
25 preparation was assessed using the gel band-shift assay and related to the total protein content of the extract as a method of standardization. The addition of herring sperm DNA depended on the purity of UL9 used in the experiment. Binding assays were incubated at 25°C
30 for 5-30 minutes.

2. Full Length UL9 Protein from the Baculovirus System.

The binding reaction conditions for the full
35 length baculovirus-produced UL9 polypeptide have also

been optimized. The optimal conditions for the current assay were determined to be as follows: 20 mM Hepes; 100 mM NaCl; 0.5 mM dithiothreitol; 1 mM EDTA; 5% glycerol; from 0 to 10^4 -fold excess of sheared herring sperm DNA; 0.005 - 0.1 ng radiolabeled (specific activity, approximately 10^8 cpm/ μ g) or digoxigenated, biotinylated oligonucleotide probe, and 5-10 μ g crude UL9 protein preparation. The full length protein also binds well under the optimized conditions established for the truncated UL9-COOH protein.

Example 5

The Effect of Test Sequence Variation on the Half-Life of the UL9 DNA:Protein Complex

The oligonucleotides shown in Figure 5 were radiolabelled as described above. The competition assays were performed as described in Example 4B using UL9-COOH. Radiolabelled oligonucleotides were mixed with the UL9-COOH protein in binding buffer (typical reaction: 0.1 ng oligonucleotide 32 P-DNA, 1 μ l UL9-COOH extract, 20 mM HEPES, pH 7.2, 50 mM KCl, 1 mM EDTA, and 1 mM DTT). The reactions were incubated at room temperature for 10 minutes. A zero time point sample was then taken and loaded onto an 8% polyacrylamide gel (run use TBE). One μ g of the unlabelled 17 bp competitive DNA oligonucleotide (SEQ ID NO:616) (Example 4B) was added at 5, 10, 15, 20, or 60 minutes before loading the reaction sample on the gel. The results of this analysis are shown in Figure 9: the screening sequences that flank the UL9 binding site (SEQ ID NO:605-SEQ ID NO:613) are very dissimilar but have little effect on the off-rate of UL9. Accordingly, these results show that the UL9 DNA binding protein is effective to bind to a screening sequence in duplex DNA with a binding affinity that is substantially

170

independent of test sequences placed adjacent the screening sequence. Filter binding experiments gave the same result.

5

Example 6

The Effect of Actinomycin D, Distamycin A, and Doxorubicin on UL9 Binding to the screening Sequence is Dependent on the Specific Test Sequence

Different oligonucleotides, each of which contained the screening sequence (SEQ ID NO:601) flanked on the 5' and 3' sides by a test sequence (SEQ ID NO:605 to SEQ ID NO:613), were evaluated for the effects of distamycin A, actinomycin D, and doxorubicin on UL9-COOH binding.

Binding assays were performed as described in Example 5. The oligonucleotides used in the assays are shown in Figure 5. The assay mixture was allowed to pre-equilibrate for 15 minutes at room temperature prior to the addition of drug.

A concentrated solution of Distamycin A was prepared in dH₂O and was added to the binding reactions at the following concentrations: 0, 1 μ M, 4 μ M, 16 μ M, and 40 μ M. The drug was added and incubated at room temperature for 1 hour. The reaction mixtures were then loaded on an 8% polyacrylamide gel (Example 5) and the components separated electrophoretically. Autoradiographs of these gels are shown in Figure 10A. The test sequences tested were as follows: UL9 polyT, SEQ ID NO:609; UL9 CCCG, SEQ ID NO:605; UL9 GGGC, SEQ ID NO:606; UL9 polyA, SEQ ID NO:608; and UL9 ATAT, SEQ ID NO:607. These results demonstrate that Distamycin A preferentially disrupts binding to UL9 polyT, UL9 polyA and UL9 ATAT.

A concentrated solution of Actinomycin D was prepared in dH₂O and was added to the binding reactions

at the following concentrations: 0 μ M and 50 μ M. The drug was added and incubated at room temperature for 1 hour. Equal volumes of dH₂O were added to the control samples. The reaction mixtures were then loaded on an 8% polyacrylamide gel (Example 5) and the components separated electrophoretically. Autoradiographs of these gels are shown in Figure 10B. In addition to the test sequences tested above with Distamycin A, the following test sequences were also tested with Actinomycin D: AToril, SEQ ID NO:611; oriEco2, SEQ ID NO:612, and oriEco3, SEQ ID NO:613. These results demonstrate that actinomycin D preferentially disrupts the binding of UL9 to the oligonucleotides UL9 CCCG and UL9 GGGC.

A concentrated solution of Doxorubicin was prepared in dH₂O and was added to the binding reactions at the following concentrations: 0 μ M, 15 μ M and 35 μ M. The drug was added and incubated at room temperature for 1 hour. Equal volumes of dH₂O were added to the control samples. The reaction mixtures were then loaded on an 8% polyacrylamide gel (Example 5) and the components separated electrophoretically. Autoradiographs of these gels are shown in Figure 10C. The same test sequences were tested as for Actinomycin D. These results demonstrate that Doxorubicin preferentially disrupts the binding of UL9 to the oligonucleotides UL9polyT, UL9 GGGC, oriEco2, and oriEco3. Doxorubicin appears to particularly disrupt the UL9:screening sequence interaction when the test sequence oriEco3 is used. The sequences of the test sequences for oriEco2 and oriEco3 differ by only one base: an additional T residue inserted at position 12, compare SEQ ID NO:612 and SEQ ID NO:613.

Example 7Use of the Biotin/Streptavidin Reporter SystemA. The Capture of Protein-Free DNA.

Several methods have been employed to sequester
5 unbound DNA from DNA:protein complexes.

1. Magnetic Beads.

Streptavidin-conjugated superparamagnetic
polystyrene beads (Dynabeads M-280 Streptavidin, Dynal
AS, $6-7 \times 10^8$ beads/ml) are washed in binding buffer then
10 used to capture biotinylated oligonucleotides (Example
1). The beads are added to a 15 μ l binding reaction
mixture containing binding buffer and biotinylated oli-
gonucleotide. The beads/oligonucleotide mixture is
incubated for varying lengths of time with the binding
15 mixture to determine the incubation period to maximize
capture of protein-free biotinylated oligonucleotides.
After capture of the biotinylated oligonucleotide, the
beads can be retrieved by placing the reaction tubes in
a magnetic rack (96-well plate magnets are available
20 from Dynal). The beads are then washed.

2. Agarose Beads.

Biotinylated agarose beads (immobilized D-
biotin, Pierce, Rockford, IL) are bound to avidin by
25 treating the beads with 50 μ g/ μ l avidin in binding
buffer overnight at 4°C. The beads are washed in
binding buffer and used to capture biotinylated DNA.
The beads are mixed with binding mixtures to capture
biotinylated DNA. The beads are removed by centrifuga-
30 tion or by collection on a non-binding filter disc.

For either of the above methods, quantification of
the presence of the oligonucleotide depends on the
method of labelling the oligonucleotide. If the oligo-
nucleotide is radioactively labelled: (i) the beads
35 and supernatant can be loaded onto polyacrylamide gels

to separate DNA:protein complexes from the bead:DNA complexes by electrophoresis, and autoradiography performed; (ii) the beads can be placed in scintillation fluid and counted in a scintillation counter.

5 Alternatively, presence of the oligonucleotide can be determined using a chemiluminescent or colorimetric detection system.

B. Detection of Protein-Free DNA.

10 The DNA is end-labelled with digoxigenin-11-dUTP (Example 1). The antigenic digoxigenin moiety is recognized by an antibody-enzyme conjugate, anti-digoxigenin-alkaline phosphatase (Boehringer Mannheim Indianapolis IN). The DNA/antibody-enzyme conjugate is

15 then exposed to the substrate of choice. The presence of dig-dUTP does not alter the ability of protein to bind the DNA or the ability of streptavidin to bind biotin.

20 1. Chemiluminescent Detection.

Digoxigenin-labelled oligonucleotides are detected using the chemiluminescent detection system "SOUTHERN LIGHTS" developed by Tropix, Inc. (Bedford, MA). Use of this detection system is illustrated in

25 Figures 11A and 11B. The technique can be applied to detect DNA that has been captured on either beads or filters.

Biotinylated oligonucleotides, which have terminal digoxigenin-containing residues (Example 1), are

30 captured on magnetic (Figure 11A) or agarose beads (Figure 11B) as described above. The beads are isolated and treated to block non-specific binding by incubation with I-Light blocking buffer (Tropix) for 30 minutes at room temperature. The presence of oligonucleotides is detected using alkaline phosphatase-

35

conjugated antibodies to digoxigenin. Anti-digoxigenin-alkaline phosphatase (anti-dig-AP, 1:5000 dilution of 0.75 units/ul, Boehringer Mannheim) is incubated with the sample for 30 minutes, decanted, and the sample washed with 100 mM Tris-HCl, pH 7.5, 150 mM NaCl. The sample is pre-equilibrated with 2 washes of 50 mM sodium bicarbonate, pH 9.5, 1 M MgCl₂, then incubated in the same buffer containing 0.25 mM 3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy) phenyl-1,2-dioxetane disodium salt (AMPPD) for 5 minutes at room temperature. AMPPD was developed (Tropix Inc.) as a chemiluminescent substrate for alkaline phosphatase. Upon dephosphorylation of AMPPD the resulting compound decomposes, releasing a prolonged, steady emission of light at 477 nm.

Excess liquid is removed from filters and the emission of light occurring as a result of the dephosphorylation of AMPPD by alkaline phosphatase can be measured by exposure to x-ray film or by detection in a luminometer.

In solution, the bead-DNA-anti-dig-AP is resuspended in "SOUTHERN LIGHT" assay buffer and AMPPD and measured directly in a luminometer. Large scale screening assays are performed using a 96-well plate-reading luminometer (Dynatech Laboratories, Chantilly, VA). Subpicogram quantities of DNA (10^2 to 10^3 attomoles (an attomole is 10^{-18} moles)) can be detected using the Tropix system in conjunction with the plate-reading luminometer.

30

2. Colorimetric Detection.

Standard alkaline phosphatase colorimetric substrates are also suitable for the above detection reactions. Typically substrates include 4-nitrophenyl phosphate (Boehringer Mannheim). Results of colorime-

35

175

tric assays can be evaluated in multiwell plates (as above) using a plate-reading spectrophotometer (Molecular Devices, Menlo Park CA). The use of the light emission system is more sensitive than the colorimetric systems.

Example 8

Labelling Test Oligonucleotides to Equivalent Specific Activities

10 The top strands of 256 oligonucleotides, containing all possible 4 bp sequences in the test sites flanking the UL9 recognition site, were synthesized. The oligonucleotides were composed of identical sequences except for the 4 bp sites flanking either side
15 of the UL9 recognition sequence (SEQ ID No:601). The oligonucleotides had the general sequence presented in Figure 14B (SEQ ID NO:617), where XXXX is the test sequence and N = A,G,C, or T. A 12 bp primer sequence, which is the complementary sequence to the 3'-end of
20 the test oligonucleotide, was also synthesized: the primer was designated the HSV primer and is presented as SEQ ID NO:618.

The HSV primer was used to prime second strand synthesis and to facilitate labeling the oligonucleo-
25 tides to the same specific activity. Oligonucleotide labelling was accomplished by labeling the 5' end of the HSV primer and then using the same primer to prime second strand synthesis of all 256 test oligonucleotides. The 5' end of the primer can be labeled with
30 radioisotopes such as ^{32}P , ^{33}P , or ^{35}S , or with non-radioactive detection systems such as digoxigenin or biotin as discussed in the Capture/Detection section.

Radioactive-labeling of the primer with ^{32}P is accomplished by the enzymatic transfer of a radioactive
35 phosphate from $\gamma\text{-}^{32}\text{P}\text{-ATP}$ to the 5' end of the primer

oligonucleotide using T4 polynucleotide kinase (Ausubel, et al.). For labeling 256 oligonucleotides, approximately 60 μg HSV primer was labeled as follows. The oligonucleotide was incubated for 1 hour at 37°C with 125 μl γ - ^{32}P -ATP (20 mCi total, 7000 Ci/mmol) and 600 units of T4 polynucleotide kinase in a 3 ml reaction volume containing 50 mM Tris-HCL, pH 7.5, 10 mM MgCl_2 , 10 mM spermidine, and 1.5 mM dithiothreitol (freshly prepared). To stop the reaction, EDTA was added to a final concentration of 20 mM. Unincorporated nucleotides were removed using "G-25 SEPHADEX" chromatography in 10 mM Tris-HCL, pH 7.5, 50 mM NaCl, and 1 mM EDTA (TE+50).

The radioactive primer was individually annealed to the top strand of each of the 256 test oligonucleotides. The bottom strand is synthesized using deoxyribonucleotides and Klenow fragment or T4 polymerase (Ausubel, et al.). The annealing mixture typically contained 200 ng HSV primer mixed with 1 μg top strand in 20 mM Tris-HCL, pH 7.5, 1 mM spermidine, and 0.1 mM EDTA (35 μl reaction volume). The primer was annealed to the top strand by incubating the sample for 2 minutes at 70°C, then placing the sample at room temperature or on ice. To the annealing mixture, 4.5 μl 10x Klenow buffer (10X = 200 mM Tris-HCL, 500 mM NaCl, 50 mM MgCl_2 , 10 mM dithiothreitol), 5 μl 0.5 mM each dNTP (dATP, dCTP, dGTP, dTTP), and 1 μl Klenow fragment were added. This reaction mixture was incubated 30-60 minutes at room temperature (or up to 37°C).

The volume of the reaction mixture was increased by adding 75 μl a solution of 10 mM Tris-HCL, pH 7.5, 50 mM NaCl, and 10 mM EDTA. The reaction mixture was applied to a 1 ml "G-25 SEPHADEX" (in TE+50) spin column. The spin columns were prepared by plugging 1cc

tuberculin syringes with silanized glass wool and adding a slurry of "G-25 SEPHADEX." The columns were prespun at 2000 rpm in a tabletop centrifuge for 4 minutes. The samples (reaction mixtures) were passed
5 through the column by centrifugation (2000 rpm, 4 minutes at room temperature) to remove unincorporated deoxyribonucleotides. The incorporation of ^{32}P was measured by placing a small volume of the sample in scintillation fluor and determining the disintegrations
10 per minute (dpms) in a scintillation counter.

The radiolabeled double-stranded oligonucleotides were then diluted to the same specific activity (equal dpms per volume). Typically, a concentration of 0.1 to 1 ng/ μl oligonucleotide was used in the assay.

15 The same procedure can be used for second strand synthesis and labeling to equal specific activity regardless of the type of label on the HSV primer.

Example 9

20 An Arrayed Sample Format

Screening large numbers of test molecules or test sequences is most easily accomplished in an arrayed sample format, for example, a 96-well plate format. Such formats are readily amenable to automation using
25 robotics systems. Several different types of disposable plastic plates are available for use in screening assays including the following: polyvinyl chloride (PVC), polypropylene (PP), polyethylene (PE), and polystyrene (PS) plates. Plates, or any testing
30 vehicle in which the assay is performed, are tested for protein and DNA adsorption and coated with a blocking reagent if necessary.

One method for testing protein or DNA adsorption to plates is to place assay mixtures in the wells of
35 the plates for varying lengths of time. Samples are

then removed from the wells and a nitrocellulose dot blot capture system (Ausubel, et al.; Schleicher and Schuell) is used to measure the amount of DNA:protein complex remaining in the mixture over time.

5 When radiolabeled oligonucleotides are used for the test, signal can be measured using autoradiography and a scanning laser densitometer. A decrease in the amount of DNA:protein complex in the absence of competitor molecules is indicative of plate adsorption.
10 If plate adsorption occurs, the plates are coated with a blocking agent prior to use in the assay.

 None of the plates listed above showed marked adsorption at a 30 minute time point under the conditions of the assay. However, most plates, regardless
15 of brand, showed significant adsorption at times greater than 2 hours.

 Coating the plates with a blocking agent decreases variability in the assay. Several types of blocking reagents typically used to block the adsorption of
20 macromolecules to plastic are known, primarily from immunoscreening procedures. For example, plates may be blocked with either 1% bovine serum albumin (BSA) in phosphate-buffered saline (PBS), or 0.1% gelatin, 0.05% "TWEEN29" in PBS.

25 To test for the effectiveness of using such blocking reagents, the plates were treated with the above reagents for 1 hour at room temperature, then washed three times with 0.05% "TWEEN20" in PBS and once with the assay buffer. Assay reaction mixtures were
30 aliquoted to the plates and tested as described above using dot blot capture assays. Both of the blocking reagents (BSA or gelatin) were effective in blocking DNA and protein binding -- except when polypropylene plates were used. Based on these experiments, PVC

plates blocked with BSA were determined to work well in the assay of the present invention.

Plates were tested for inter- and intra-plate variability by aliquoting duplicate samples to all 96-wells of several plates, and determining the amount of DNA:protein complex recovered using the dot blot/nitrocellulose system. The coefficient of variation [%CV = (the standard deviation/mean)*100] was calculated for intra-plate variability (i.e., between samples on the same plate) and inter-plate variability (i.e., between plates). Blocked PVC plates showed an intra-plate %CV of 5-20%; inter-plate variability was about 8%.

15

Example 10

Sequence Selectivity and Relative Binding Affinity for Distamycin

Using the assay method of the present invention, distamycin was tested for sequence selectivity and relative binding affinity to 256 different 4 bp sequences.

20

A. The Assay Mixture.

Water, buffer and UL9 were mixed on ice and aliquoted to the wells of a 96-well plate. The addition of water/UL9/buffer mix was accomplished with an 8-channel repipettor, which holds a relatively large volume and allowed rapid, accurate pipetting to all 96 wells of a master experimental plate.

30

Radiolabeled double-stranded oligonucleotides were aliquoted from 96-well master stock plates (containing the array of all 256 oligonucleotides diluted to the same specific activity) to the wells of the master experimental plates.

Master assay mixtures in the master experimental plates were thoroughly mixed by pipetting up and down. The mixtures were aliquoted to the test plates. Each test plate typically included one sample as a control
5 (no test molecules added) and as many test samples as were needed for different test molecules or test molecule concentrations. There were 3 master oligonucleotide stock plates, containing the array of 256 oligonucleotides. Accordingly, an experiment testing distamycin
10 at different concentrations would require 256 control assays (one for each oligonucleotide) and 256 assays at each of the drug concentrations to be tested.

The following assay mixture was used for testing distamycin in the assay of the present invention: 1.5
15 nM radiolabeled DNA and 12.8 nM UL9-COOH protein (prepared as described above in the UL9 binding buffer; 20 mM Hepes, pH 7.2, 50 mM KCl, and 1 mM dithiothreitol). The concentration of the components in the assay mixture can be varied as described above in the
20 Detailed Description.

Assay mixtures containing both UL9 and DNA were incubated at room temperature for at least 10 minutes to allow the DNA:protein complexes to form and for the system to come to equilibrium. At time = 0, the assay
25 was begun by adding water (control samples) or distamycin (5-15 μ M, test samples) to the assay mixtures using a 12-channel micropipettor. After incubation with drug for 5-120 minutes, samples were taken and applied to nitrocellulose on a 96-well dot blot apparatus (Schleicher and Schuell). The samples were held at 4°C.
30

Tests were performed in duplicate. Typically, one set of 256 test oligonucleotides was scrambled with respect to location on the 96-well plate to eliminate any effects of plate location.

B. The Capture/Detection System.

A 96-well dot blot apparatus was used to capture the DNA:protein complexes on a nitrocellulose filter. The filters used in the dot blot apparatus were pretreated as follows. The nitrocellulose filter was pre-wetted with water and soaked in UL9 binding buffer. The filter was then placed on 1 to 3 pieces of 3MM filter paper, which were also presoaked in UL9 binding buffer. All filters were chilled to 4°C prior to placement in the apparatus.

Prior to the application of the assay sample to the wells of the dot-blot apparatus, the wells were filled with 375 µl of UL9 binding buffer. Typically, 5-50 µl of sample (usually 10-15 µl) were pipetted into the wells containing binding buffer and a vacuum applied to the system to pull the sample through the nitrocellulose. Unbound DNA passes through the nitrocellulose, protein-bound DNA sticks to the nitrocellulose. The filters were dried and exposed to X-ray film to generate autoradiographs.

C. Quantitation of Data.

The autoradiographs of the nitrocellulose filters were analyzed with a Molecular Dynamics (Sunnyvale, CA) scanning laser densitometer using an ImageQuant software package (Molecular Dynamics). Using this software, a 96-well grid was placed on the image of the autoradiograph and the densitometer calculated the "volume" of each dot ("volume" is equivalent to the density of each pixel in the grid square multiplied by the area of the grid square). The program automatically subtracts background. The background was determined by either the background of a line or object drawn outside the grid or by using the gridlines as background for each individual dot.

The data is exported to a spreadsheet program, such as "EXCEL" (Microsoft Corporation, Redmond, WA) for further analysis.

5 D. Analysis of Data.

The data generate from the densitometry analysis was analyzed using the spreadsheet program "EXCEL."

For each test oligonucleotide, at each drug concentration and/or each time point, a raw % score was
10 calculated. The raw % score (r%) can be described as

$$r\% = (T/C) \times 100$$

where T was the densitometry volume of the test sample
15 and C was the densitometry volume of the control sample. The oligonucleotides were then ranked from 1 to 256 based on their r% score. Further calculations were based on the rank of each oligonucleotide with respect to all other oligonucleotides.

20 The rank of each oligonucleotide was averaged over several experiments (where one experiment is equivalent to testing all 256 test oligonucleotides by the assay of the present invention) in view of the variability in rank between any two experiments. The confidence level
25 for the ranking of the oligonucleotides increased with repetition of the experiment.

Figure 15 shows the results of 4 separate experiments with distamycin. The test samples were treated with 10 μ M distamycin for 30 minutes. The r% scores
30 are shown for each of the 4 experiments (labeled 918A, 918B, 1022A, and 1022B) and the ranks of each oligonucleotide in each experiment are shown. The test oligonucleotides have been ranked from 1 to 256 based on their average rank. The average rank was the sum of

the ranks in the individual experiments divided by the number of experiments.

Figures 16 and 17 show the results presented in Figure 15 in graphic form. Figure 16 shows the average ranks plotted against the ideal ranks 1 to 256. Figure 17 shows the average r% scores plotted against the rank of 1 to 256. These data demonstrate the reproducible ability of the assay to detect differential binding and effects of distamycin on different 4 bp sequences.

Example 11

Determining a Consensus Binding Site for Distamycin

One method used to determine the sequence preferences for distamycin was to examine the sequences that rank highest in the assay for sequence similarities. This process may be accomplished visually or by designing computer programs to inspect the data.

Using the data shown in Figure 15, consensus sequences can be constructed for distamycin in the following manner. Sequences with rankings less than 50 (indicating a strong effect of distamycin on the test sequence) in all four experiments were:

TABLE VI

Sequence	Rank
TTCC	1
TTAC	2
TACC	3
TATC	4
TTCG	6
ACGG	8

Sequences with rankings less than 50 (indicating a strong effect of distamycin on the test sequence) in three of the four experiments were:

5

TABLE VII

10

15

20

Sequence	Rank
AACG	5
TTTC	7
TTAG	10
TAAC	12
TACG	15
AGAC	17
AAAC	18
AGCG	21
AGCC	22
TTCT	24
ACGC	25
AGGG	28
AGGC	30
TTGC	37
ATCG	39
TTTG	43

Sequences with rankings less than 50 (indicating a strong effect of distamycin on the test sequence) in two of the four experiments were:

25

TABLE VIII

30

Sequence	Rank
TAGC	9
TTGG	11

185

Sequence	Rank
AAAG	13
AACC	14
CAAC	16
ATCC	19
AAGG	20
TAAG	23
ACCC	26
TCCC	29
TATG	31
ACCG	32
TCGG	34
AGTC	35
CTCG	38
AATC	44
AGAG	46
TTAA	47
ACAC	48
AGTG	49
TCAC	52

The following assumptions allow prediction of a consensus sequence for a distamycin recognition sequence: (i) the most favored sequences are the test sequences that rank in the top 50 in all four experiments; (ii) the next favored sequences will be the test sequences that rank in the top 50 in 3 of 4 experiments; and (iii) the next favored sequences will be the test sequences that rank in the top 50 in 2 of 4 experiments.

186

The positions in the test sequence are represented by the numerals 1, 2, 3 and 4. One consensus sequence that predicted from the above binding data is:

5 1 2 3 4
 T T/A N C/G

The nucleotides at each position can also be ranked:

 1 2 3 4
 T T>A C>A>T>G C>G

10 Furthermore, the importance of the position of the nucleotide can be ranked. Examination of this data would indicate that the importance of the positions is
 1 > 4 > 2 > 3.

15 These data can be tested for validity by deriving all possible consensus sequences and examining their scores in the assay. The consensus sequences derived from the above information, in order of rank as predicted by the consensus sequence, are:

20

TABLE IX

25

30

Sequence	Predicted Rank	Actual Rank
TTCC	1	1
TACC	2	3
TTCG	3	6
TACG	4	15
TTAC	5	2
TAAC	6	12
TTAG	7	10
TAAG	8	23
TTTC	9	7
TATC	10	4
TTTG	11	43
TATG	12	31

187

Sequence	Predicted Rank	Actual Rank
TTGC	13	37
TAGC	14	9
TTGG	15	11
TAGG	16	<u>58</u>
	Average rank:	17

5

Note that the actual rank numbers are out of a possible 256 and that only one number is greater than 50. The average rank of these 16 oligos is only 17. These data indicate that the consensus sequence has predictive value.

10

Using the same data, a second consensus sequence can be derived that has slightly worse average rank with respect to the relative effect of distamycin in the assay.

15

TABLE X

1	2	3	4
A	A/G/C	G/C/A	G/C
	A>G=C	C>A=G	G=C

20

The test sequences predicted by this consensus sequence are as follows:

TABLE XI

Sequence	Actual rank
AACG	5
AACC	14
AAAG	13
AAAC	18
AAGG	20

25

30

188

Sequence	Actual rank
AAGC	74
AGCG	21
AGCC	22
AGAG	46
AGAC	17
AGGG	28
AGGC	30
ACCG	32
ACCC	26
ACAG	73
ACAC	48
ACGG	8
ACGC	<u>25</u>
Ave. rank:	29

This consensus sequence also appears to be predictive of favored distamycin binding sites since the average rank of test oligonucleotides predicted by this sequence is 29, substantially below the median rank of 128. However, the sequences predicted by this consensus sequence do not appear to be affected as strongly by distamycin as the sequences in the first consensus sequence, described above.

25

Example 12Testing Actinomycin D to Determine Sequence Specificity and Relative Binding AffinityA. Ranking of Actinomycin D Sequence Binding Affinities.

Actinomycin D has been tested for sequence selectivity and relative binding affinity to the 256 different 4 bp sequences. The assay was performed essentially as described in Example 10. One assay mixture useful for the testing of actinomycin D contained 1.5 nM radiolabeled DNA and 12.8 nM UL9-COOH protein prepared as described above in the UL9 binding buffer (20 mM Hepes, pH 7.2, 50 mM KCl, and 1 mM dithiothreitol). The concentration of the components can be varied as described in the Detailed Description.

The assay mixtures containing both UL9 and DNA were incubated at room temperature for at least 10 minutes to allow the DNA:protein complexes to form and for the system to come to equilibrium. At time = 0, the assay was begun by adding water (control samples) or actinomycin D (25 μ M, test samples) to the assay mixtures using a 12-channel micropipettor. After incubation with drug for 30 minutes, samples were taken and applied to nitrocellulose filters using a 96-well dot blot apparatus (Schleicher and Schuell) held at 4°C. Figure 18 shows the results of 8 screens of actinomycin D.

The % reduction in DNA:protein complex as a result of the presence of actinomycin D is called "r%"; the lower the r% score, the more effective the test molecule in blocking the DNA:protein interaction. For each screen, the test oligonucleotides have been ranked from 1 to 256, based on the r% score; the rank of 1 denotes the lowest r% score (the test oligonucleotide most effected by the test molecule), the rank of 256 denotes

the highest r% score (the test oligonucleotide least effected by the test molecule). The table also shows the average r% score and average rank of each test oligonucleotide; the averages are calculated from the sum of the individual scores and ranks divided by the number of screens, respectively. The test oligonucleotides are then ranked from 1 to 256 based on the average rank in all screens. The final ranking is shown in the two external columns on the table. Test oligonucleotides ranking less than 50 in any individual screen are shown in highlighted boxes.

Figure 19 shows the final rank of test oligonucleotides screened with actinomycin D plotted against the average r% score for these test oligonucleotides.

Figure 20 shows the final ranking vs. the ranks in each individual experiment, the average rank, and the ideal rank.

B. Analysis of the Data Obtained from Ranking Actinomycin D Sequence Binding Affinities.

Several simple analytical procedures may be applied to the data from the screens.

1. Position Effects.

First, to examine possible preferences of the test molecule for a base at any particular position in the test site, the average r% scores are examined. The average r% scores for each of the 64 possible test oligonucleotides at each position in the test site are averaged. For example, to determine the effect of having an A in the first position of the test site, the "A," position, the average r% scores for the 64 test oligonucleotides with A in the first position are averaged. The results of this analysis are shown in Figure 21. The mean score for all oligonucleotides in

these screens was r% value 67; the standard deviation was 11.8.

If the r% score is expressed as variance from the mean, as shown in Figure 21, one observes that none of the scores is markedly deviant from the mean. These results suggest that a single base in any particular position has little impact on the binding of the actinomycin D to the test site.

2. Dinucleotide Analysis.

The results of the actinomycin D screen were examined for the presence of dinucleotide pairs that scored well or poorly in the rankings. High scores indicate a preference for the test sequence. Low scores indicate a repulsion of actinomycin D for the test sequence. A dinucleotide analysis is one of many simple analytical procedures that may be applied to the data to extract meaningful impressions about the nature of the sequences to which the test molecule has high affinity.

The data are examined in a manner similar to that used for the single nucleotide analysis. The 16 possible average r% scores for any particular dinucleotide combination are examined. Specific adjacent dinucleotides (N_1N_2 , N_2N_3 , N_3N_4) or adjacent dinucleotide pairs at any particular position (N_xN_{x+1} = the average of N_1N_2 , N_2N_3 , and N_3N_4) may be examined, as well as specific dinucleotide pairs that are not adjacent (N_1N_3 , N_2N_4 , N_1N_4) and any dinucleotide pair separated by one base (N_xN_{x+2} = the average of N_1N_3 and N_2N_4). The means for each set are determined as well as standard deviations.

The difference from the mean (i.e., the mean score less the average r% score for any particular dinucleotide) reflects the extent of deviation from the norm.

Differences from the mean greater than 2-3 standard deviations from the mean are considered to be significant. The data for the dinucleotide analysis of actinomycin D is shown in Figure 22. The differences from the mean are displayed graphically in Figure 23.

In reference to Figures 22 and 23, the dinucleotide preference of actinomycin D is GC, particularly in the M_1N_2 position, but also at any (N_iN_{i+1}) adjacent dinucleotide sequence in the test site.

If the data are combined in a combined bar chart, shown in Figure 24, where the cumulative results for any dinucleotide pair are tabulated in a single bar, the overall observation can be made that actinomycin D prefers GC-rich sequences over AT-rich sequences, with a particular preference for the dinucleotide pairs involving GC.

Example 13

A Method for Selecting Target Sites for DNA-Binding Molecules that are Dimers or Trimers of Distamycin

Once the relative binding preferences of a distamycin have been determined, sequences are selected for target sites for DNA-binding molecules composed of two distamycin molecules, bis-distamycins, or three distamycin molecules, tris-distamycins.

A. Selecting Sequences for Binding with Highest Affinity to Distamycin Oligomers.

The top binding sites for distamycin, determined as described above, are defined by the consensus sequence, 5'-T:T/A:C/A:C-3': accordingly, the top sequences are TTCC, TTAC, TACC and TAAC. Using this information, $2^4 = 16$ possible dimer sequences, i.e., combinations of the four top binding sequences, can be

targeted by a bis-distamycin in which the distamycin molecules are immediately adjacent to one another.

The top strands of the 16 possible duplex DNA target sites for binding bis-distamycins are shown in Figure 25. Similarly, trimers of distamycin, tris-distamycins, could be targeted toward selected 12 bp sequences, comprised of all possible combinations of the four 4 bp sequences. There are $3^4 = 81$ possible highest affinity target trimer sequences.

There are several advantages to targeting longer sequences with bis- or tris-distamycin:

B. As the Number of Potential Target Sites Decreases, Specificity Increases.

All 8 bp combinatorial possibilities of the 4 top favored binding sites for distamycin are potential high affinity binding sites for bis-distamycin. The consensus sequence used in this example predicts four favored binding sites for distamycin. This represents $(4/4^4) \times 100 =$ about 1.6% of the possible 4 bp sites in the genome. Since there are 4^8 possible 8 bp sequences, this represents, on average, only $(2^4/4^8) \times 100 =$ about 0.02% of the total genome. There are 4^{12} possible 12 bp sequences, this represents, on average, only $(3^4/4^{12}) \times 100 = 0.00000075\%$ of the genome.

The following discussion provides perspective and illustrates the improvement in the actual number of target sites in the human genome for when using a dimer of distamycin versus a monomer of distamycin. The human genome is about 3×10^9 bp. If the number of favored target sites for distamycin is four, and the number of possible 4 bp sequences is $4^4 = 256$, then the number of favored target sites in the genome is $(4/256)(3 \times 10^9) = 4.7 \times 10^7$, or about 50 million favored target sites.

Given that the number of possible 8 bp sites is $4^8 = 65,536$, if all possible combinatorial 8 bp sites derived from the favored 4 bp sites ($2^4 = 16$; Figure 25) are favored, then the number of favored 8 bp target sites is $(16/65,536)(3 \times 10^9) = 7.3 \times 10^5$ or about 700,000 possible sites. This represents a 64-fold reduction in the number of highest affinity target sites between distamycin and bis-distamycin; alternatively, this result can be viewed as a 64-fold increase in specificity.

Likewise, given that the number of possible 12 bp sites is $4^{12} = 1.7 \times 10^7$, if all possible favored 12 bp sites ($3^4 = 81$) are favored, then the number of favored 12 bp target sites is $(81/1.7 \times 10^7)(3 \times 10^9) = 1.4 \times 10^4$: i.e., 14,000 possible highest affinity sites. This represents an approximately 3000-fold decrease in the number of highest affinity target sites between distamycin and tris-distamycin and a 500-fold decrease in the number of highest affinity target sites between bis-distamycin and tris-distamycin.

C. An Exponential Increase in Affinity.

As the target site increases in size, (i) the number of target sites in a defined number of nucleotides decreases, and (ii) the specificity increases. Further, the affinity of binding is typically the product of the binding affinities of component parts (see Section VI.E.1 above). As an example, the published binding constant for distamycin to bulk genomic DNA is about $2 \times 10^5 \text{ M}^{-1}$. Dimers of distamycin will have a theoretical binding affinity of the square of the binding constant of distamycin:

$$(K_{\text{dista,average}} = 2 \times 10^5 \text{ M}^{-1}; K_{\text{bis-dista}} = (2 \times 10^5 \text{ M}^{-1})^2 = 4 \times 10^{10} \text{ M}^{-1}).$$

Trimers of distamycin will have binding

195

affinities of the cube of the binding affinity of distamycin:

$$(K_{\text{tris-dista}} = (2 \times 10^5 \text{ M}^{-1})^3 = 8 \times 10^{15} \text{ M}^{-1}).$$

5

Thus, if distamycin shows only a 10-fold higher affinity ($2 \times 10^6 \text{ M}^{-1}$) for the top favored binding sites than the average binding sites in DNA, then the affinity constant for bis-distamycin to an 8 bp site comprised of two favored binding sites is 100-fold higher than for an 8 bp sequence comprised of two average binding sites:

$(K_{\text{bis-dista, favored sites}} / K_{\text{bis-dista, average sites}} = (2 \times 10^6)^2 / (2 \times 10^5)^2 = 100).$ While this does not represent absolute sequence specificity in binding, the binding affinity is 100-fold greater for 0.02% (16/65,536) of the total possible 8 bp target sequences.

The use of a trimer targeted sequence will afford an even higher increase in affinity to the most favored binding sites:

$K_{\text{tris-dista, favored sites}} / K_{\text{tris-dista, average sites}} = (2 \times 10^6)^3 / (2 \times 10^5)^3 = 1000.$ Thus, with only 10-fold differential activity in binding between favored sites and average sites, a 1000-fold difference in affinity can be achieved by designing trimer molecules to specific target sites. When considering the administration of DNA-binding molecules as drugs, a 1000-fold lower dose of tris-distamycin, versus the distamycin monomer, could be administered and an increase in relatively specific binding to selected target sites achieved.

In this example, the differential activity of distamycin is only 10-fold. Clearly, differential activities of larger magnitudes will greatly accentuate the increased affinity effect. For example, a 100-fold

difference in activity of a 4 bp DNA-binding molecule toward high affinity and average affinity sequences would result in (i) a 10,000-fold difference in the binding affinity of a dimer of the molecule targeted to an 8 bp sequence, and (ii) a million-fold increase in the binding affinity of the trimer to a 12 bp sequence.

D. Selecting Target Sequences for Distamycin Oligomers with Flexible and/or Variable-Length Linkers in Between the Distamycin Moieties.

The sequences that can be targeted with bis- or tris-distamycin molecules are not limited to sequences in which the two 4 bp favored binding sites are immediately adjacent to one another. Flexible linkers can be placed between the distamycin moieties and sequences can be targeted that are not immediately adjacent. The target sequences can have distances of 1 to several bases between them: this distance depends on the length of the chemical linker. Examples of bis-distamycin target sequences for bis-distamycins with internal flexible and/or variable length linkers targeted to sites comprised of two TTCC sequences are shown in Figure 26, where N is any base.

For each particular bis-distamycin, the explanations of increased affinity and specificity remain the same as described above with the following exception. For the case in which the linker was sufficiently flexible to span different numbers of bases in between the two distamycin sites, the number of sites targeted with highest affinity would be multiplied by the number of bases spanned.

In respect to the ease of drug design and target selection, there are several advantages to the above described targeting strategies, including the following:

i) Any conformational changes induced by binding at the half-site would be minimized.

ii) The affinity, therefore, would be more likely to be the product of the affinities of the interactions observed for the monomeric sites.

iii) The half-molecule (e.g., 1 distamycin unit) would anchor the bis-molecule (e.g., bis-distamycin) thus increasing the localized concentration for the binding of the second half of the bis-molecule.

iv) If a simple linking chain is used, with a variable number of atoms, the number of sites that can be targeted by multimers of the monomer increases. This targeting method can be of value when, for example, there are no medically significant target sites with adjacent favored binding sites for distamycin. Therefore there are no good target sites for bis-distamycin. In this situation, the database can be screened for additional target sequences with N_1 to N_n (where N is any base) between the two target binding sequences. For example, where $n=4$, the number of sequences to be searched becomes $(4^2)*4 = 64$. The likelihood of finding such a sequence is reasonably high.

E. Selecting a Specific Target Site.

Using the above approach, a sequence was identified from the medically significant target site database that contains SEQ ID NO:619, which is a subset of the group of sequences represented by SEQ ID NO:620. SEQ ID NO:619 occurs overlapping the binding site for a transcription factor, Nuclear Factor of Activated T Cells (NFAT-1), which is a major regulatory factor in the induction of interleukin 2 expression early in the T cell activation response. NFAT-1 is crucial in (i) the T cell response, and (ii) in blocking the expres-

sion of IL-2, which causes immunosuppression. The sequences TTCC and TTTC, the distamycin target binding sequences in SEQ ID NO:619, rank first and seventh in the assay.

5

Example 14

The Use of the Assay in Competition Studies

The assay of the present invention measures the effect of the binding of a DNA-binding molecule to a test site by the release of a protein from an adjacent screening site. Accordingly, the assay is an indirect assay. Following here is the description of an application of the assay useful to provide confirmatory evidence of the data obtained in the initial screening processes.

15

The results of the distamycin screening assay described in Example 10 suggested that there were possible false negatives: specifically, test sequences that bind distamycin but fail to show an effect on the binding of the reporter protein. The data suggesting false negatives was as follows. If the assay detected strictly the affinity of binding of distamycin, then the scores of the test sequences complementary to the high-scoring test sequences should always be equally high. However, an examination of the highest ranking test sequences and the complementary test sequences reveals that this is not the case (see Table XII).

20

25

TABLE XII

30

Rank	Test Sequence	Complement	Rank of complement
1	TTCC	GGAA	42
2	TTAC	GTAA	244
3	TACC	GGTA	185

Rank	Test Sequence	Complement	Rank of complement
4	TATC	GATA	213
5	AACG	CGTT	144
6	TTCG	CGAA	216
7	TTTC	GAAA	235

5 All but one of the complementary sequences rank in the lower half, 4 of them in the lowest 20%, *i.e.*, these was little effect on reporter protein binding in the presence of distamycin when using these sequences
10 as test sequences in the assay.

This observation reflects the usefulness of a confirmatory assay that examines the relative affinity of a particular sequence for binding distamycin. A confirmatory assay may also be useful in revealing
15 additional information about the physical characteristics of drug binding. For example, one can hypothesize that the reason for the apparent inverse relationship between test sequences with high activity in the assay and their complements is that the effect of distamycin
20 is directional and only active at one test site. This hypothesis can be tested using the following competition experiment. Competitor oligonucleotides, containing test sequences of interest, are added to the assay mixture. This allows the determination of which test
25 sequences compete most effectively with the radiolabelled test oligonucleotide for binding distamycin.

Assay mixtures are prepared as described in Example 10, using a high-ranking test oligonucleotide, *e.g.*, TTCC (ranking = #1), as the radiolabelled oligo-
30 nucleotide in the experiment. The test oligonucleotide TTCC is labelled to high specific activity with γ -³²P-ATP as described in Example 8: in this example, the

200

labeled TTCC oligonucleotide will be referred to as the "high specific activity test oligonucleotide".

The competitor oligonucleotides are labeled as described in Example 8, except that the ATP used for
5 kinasing the primer is 1:200 radiolabeled:nonradiolabeled. In other words, the competitor oligonucleotides are tracer labeled with radioactive phosphorous to a 200-fold lower specific activity than the high specific activity test oligonucleotide. Since all of the
10 competitor oligonucleotides are labeled with the same radiolabeled primer molecule, the relative concentrations of the competitor DNAs can be determined with high accuracy. Further, since the specific activity is the same, the concentrations can be adjusted to be the
15 same. For the purposes of this example, the competitor DNAs are referred to as "low specific activity competitor oligonucleotides."

The use of competitor DNAs for which the concentration is known is important for the competition
20 experiment. The accuracy of the competition assay may be further enhanced by separating any unincorporated radiolabeled primer from the double stranded competitor oligonucleotides. This separation can be achieved using, for example, a 6-20% polyacrylamide gel. The
25 gel is then exposed to x-ray film and the amount of double-stranded oligonucleotide determined by use of a scanning laser densitometer, essentially as described in the Examples above.

The competition assay is performed as described in
30 Example 10, except that competitor DNAs are added in increasing relative concentration to the high specific activity test oligonucleotide. The DNA concentration ([DNA]) is held constant and the UL9 concentration ([UL9]) and distamycin concentration ([distamycin]) are

201

as described in Example 10. The components in the competition assay samples are as follows.

Controls:

5 UL9 + TTCC*; UL9 + TTCC* + Competitors; UL9
 + TTCC* + distamycin;

Test samples:

 UL9 + TTCC* + distamycin + Competitors;
where UL9 is UL9-COOH, TTCC* is the high specific
activity test oligonucleotide, and Competitors are the
10 low specific activity competitor oligonucleotides.

 TTCC-low (the tracer-labeled low specific activity
competitor) competes with TTCC* on an equimolar basis
for the binding of both protein and distamycin. A com-
petitor molecule with lower affinity for distamycin
15 than TTCC requires a higher molar ratio to TTCC* to
compete for distamycin binding. The competition for
protein between all competitors is constant. Only the
competition for distamycin varies; the variability is
due to the differential affinity of the competitor oli-
20 gonucleotides for distamycin. The concentration of
competitor used in these experiments varies over a
range of concentrations and is determined empirically
by (a) the test molecule concentration, and (b) the
relative affinity of the competitor and the radio-
25 labeled test oligonucleotide. Typically, the competi-
tor DNA consists of only the test sequence, that is, no
additional sequences are connected to the test se-
quence.

 The competition assay described here facilitates
30 the determination of actual rank between the test oli-
gonucleotides that are detected as highly effective
molecules in the original assay. The competition assay
also facilitates the detection of false negatives. As
described above, the results of the assay discussed in
35 Example 10 imply "directional" binding of distamycin,

in which the effect of binding is only detected when the molecule is bound in one direction with respect to the UL9 protein. Binding in the opposite direction (i.e., to the complementary test sequence) is not
5 detected with the same activity in the assay.

The purpose of this competition experiment is to use the test oligonucleotides to compete for the binding of distamycin. If the sequences complementary to the "best binders" are false negatives in the assay,
10 they should nonetheless be effective competitors in the competition assay.

Example 15

A Method of Selecting Target Sequences 15 From Database Sequence Information

The binding of a drug or other DNA-binding molecule to the recognition sequence for TFIID, or other selected transcription factors, is expected to alter the transcriptional activity of the associated gene.
20 TATA-boxes, which are the recognition sequences for the transcriptional regulatory factor TFIID, are associated with most eukaryotic promoters and are critical for the expression of most eukaryotic genes. Targeting a DNA-binding drug to TATA boxes in general would be undesirable.
25 However, sequences flanking TATA box sequences are typically unique between genes. By targeting such flanking sequences, perhaps with one base overlapping the TFIID recognition site, each gene can be targeted with specificity using the novel DNA-binding molecules
30 designed from the data generated from the DNA-binding drug assay. One method for determining novel and specific target sequences for novel DNA-binding drugs is described here. The method may be applied to any known binding site for any specific transcription

factor, regardless of whether the identity of the transcription factor itself is known.

TATA-boxes have been determined for a large number of genes. Typically, the TATA-box consensus sequence has been identified by examining the DNA sequence 5' of the RNA start site of a selected gene. However, the most rigorous determinations of TATA boxes have also demonstrated the transcription factor binding site by DNA protection experiments and DNA:protein binding assays (using electrophoretic methods). Many of these sites are annotated in the public databases "EMBL" and "GENBANK", which both contain sequences of nucleic acids sequences. Unfortunately, the flat field listing of these databases do not consistently annotate these sites. It is possible, however, to automatically search a database, using a text parsing language called AWK, to extract most sequence information that relates to annotated promoter sequences.

The following is a description of how selected promoter sites were located in the public database from "EMBL." The flat field annotations from "EMBL" Version 32 as processed by "INTELLIGENTICS" (Mountain View, CA), were obtained with the set of UNIX programs call "IG-SUITE." These programs were executed on a "SUN IPX" workstation. An AWK script was used to parse all the primate annotation files listed in the "EMBL" database. The AWK interpreter is supplied as part of the system software that comes with the "SUN IPX" workstation.

The following is a description of how the AWK parses annotation files looking for and printing information relating to promoters and TATA-boxes. The system is asked to examine the input files for certain key words in the header lines or annotations to the sequence. The AWK interpreter reads input files line by

line and executes functions based on patterns found in each line. In this case, the AWK system read the annotation files of EMBL. The following is a description of how the AWK script can be used to parse out sequences containing TATA-boxes.

The program first examines the files for all header lines containing the word "complete" but not "mRNA" or "pseudogene"; the output is printed. Complete genes sometimes contain the promoter sequences but complete mRNA genes do not contain the promoters. mRNA genes are not of interest for the purpose of detecting promoter elements. Next, the AWK system looks for the word "exon 1" and if it finds it prints the header and "DE" line. Then it looks for "5'" and prints the header line if it does not contain the word "mRNA". Next it looks for the word "transcription" and if it finds it prints the preceding and following line along with description line.

Next, the AWK system examines the files for the word "TATA" in the header lines or references. This results is printed. After this it looks for the word "promoter" and if it finds it prints that line and the line after it which contains the information about the promoter. Then the program looks for "protein_bind" and prints that line along with the next one. The description of "protein_bind" is usually used to mark potential binding sites of transcription factors in the "EMBL" database. AWK then scans for any annotated primary mRNA start sites. The promoter sequence is found in front of the start site. Finally, any exon 1 start sites that are annotated in the feature table are extracted. Exon 1 start sites should also be the primary transcription start site and the TATA boxes usually are found approximately 25-35 base pairs 5' to the transcriptional start site.

205

The actual AWK script is included here as an example of how to parse a database to extract promoter sites:

```

BEGIN {print_next_line=0}
5  {if (print_next_line==1)
    {print $0
     print_next_line=0}
   }
   {if ($0 ~/^>/)
10      { Locus=$0
        l_flag=0 }
     }
     /^>/ && / [Cc]omplete/ && $0 !~ /mRNA|mrna/ && $0
     !~/pseudogene/{print}
15  /^>/ && /exon 1[^0-9]/ {print}
     /^>/ && /5'/ && $0 !~ /mRNA|mrna/ {print}
     /[Tt]ranscription/ {print Locus "\n" PL "\n"
$0;print_next_line=1}
     {if ($0 ~/^FT/ && $0 ~/-TATA/ && $0 ~/-note/)
20      {print Locus "\n" PL"\n"$0}
     }
     {if ($0 ~/^FT/ && $0 ~/[Tt]ranscription/ && $0 ~/\///)
     {print Locus "\n" PL"\n"$0}
     }
25  {if($2 !~ /note/ && $2 ~ /TATA/) {print Locus "\n" $0}
   }
   {if ($2 ~/-promoter/)
     {print_next_line=1
      if(l_flag==0)
30      {print Locus "\n" $0
        l_flag=1}
      else
        print $0
      }
35  }

```

206

```

    {if ($2 ~/protein_bind/)
        {print Locus "\n" $0
         print_next_line=1}
    }
5   {if ($2 ~/prim_transcript/ && $3 !~/^1..|^<1../)
        {print Locus "\n" $0
         print_next_line=1}
    }
    {if ($0 ~/^FT/ && $0 ~/number=1[^0-9]/)
10      if(PL ~/exon/){print Locus "\n" PL"\n"$0}
    }
    {PL=$0}

```

15 After the AWK script is run on the database the output is manually examined. Those sites that are clearly promoter sites are noted and nucleotide coordinates recorded. Other gene sequences are examined using the "FINDSEQ" program of "IG_SUITE" to see if the promoter sites can be determined or if the
 20 references in the database describe the promoter sequences. If so, those nucleotide coordinates are noted. At the end of this examination "FINDSEQ" is used to extract any sequences containing promoter sequences by using an indirect file of "LOCUS" names
 25 constructed using a text editor.

A parsing program was also written to extract each of the annotated sites from the file that "FINDSEQ" extracted from "EMBL." This program extracts the following information: the promoter site name and four
 30 numbers representing the nucleotide coordinates of where the sequence is to start, what the coordinate of the first base of the site is, the coordinate of the last base of the site and the end of the sequence to be extracted. A large batch file was constructed to

automatically extract each of the promoter sites. These sequences formed the basis of Table V.

The Sequence Listing presents a number of sequences that are useful as test sequences in the present invention. SEQ ID NO:1 to SEQ ID NO:481 and SEQ ID NO:600 correspond to promoter targets (typically, TATA box-containing sites) for human genes. SEQ ID NO:482 to SEQ ID NO:599 correspond to promoter targets for viral genes.

10

Example 16

Using Normalized Values to Determine Sequence Specificity and Relative Binding Affinity

A. The Assay Mixture and Calibrator Samples.

The assay mixture is prepared as described in Example 10. The concentration of the components can be varied as described in the Detailed Description.

The assay mixtures containing both UL9 and DNA are incubated at room temperature for at least 10 minutes to allow the DNA:protein complexes to form and for the system to come to equilibrium. At time = 0, the assay is begun by adding water (control samples) or test molecule (typically at 1-5 μ M, test samples) to the assay mixtures using a 12-channel micropipettor. After incubation with drug for 5-120 minutes, samples are taken and applied to nitrocellulose filters using a 96-well dot blot apparatus (Schleicher and Schuell) held at 4°C.

Calibrator samples are used to normalize the results between plates, that is, to take plate-to-plate variability into account. Calibrator samples are prepared using 2-fold serial dilutions of DNA in the assay mixture and incubating duplicate samples in one column of the 96-well assay plate. The highest concentration of DNA used is the same concentration

used in the screening samples. In general, calibrator samples are used in all experiments. However, use of calibrator samples appears to be less important for experiments using blocked plates since the variability
5 between blocked plates is lower than between unblocked plates.

The calibrator samples are used to normalize the values between plates as follows. The volume values (Example 10) for the calibrator samples are obtained
10 from densitometry. Volume values are plotted against DNA concentration. The plots are examined to ensure linearity. The volume values for the points on the calibrator line are then averaged for each plate. A factor, designated the normalization factor, is then
15 determined for each calibrator line. When the normalization factor is multiplied by the average of the points on each calibrator line, the product is the same number for all plates. Usually, the average of the line averages is used for determining the normalization
20 factor, although in theory, any of the line average numbers can be used. The operating assumption in this analysis is that the differences in the calibrator samples reflected the differences in adsorption for each plate. By normalizing to the calibrator samples,
25 these variations are minimized.

Once the normalizing factor is obtained, all of the raw volume values for each of the test assays on the plate is multiplied by the normalizing factor. For example, if the following data were obtained, the
30 process of normalization would be as follows:

TABLE XIII

PLATE NUMBER	DNA CONCENTRATION				
	0.8	0.4	0.2	0.1	Average
Plate I:	4000	2000	1000	500	1875
Plate II:	4200	2100	1050	525	1969
Plate III:	3800	1900	950	475	1781
	Average:				1875

Plate I has a normalization factor of 1; Plate II has a normalization factor of $1875/1969 = 0.95$; Plate III has a normalization factor of $1875/1781 = 1.05$. The equation used to establish these numbers is as follows: "Average average"/line average = normalization factor.

If the normalization factors are different, these factors are incorporated into the data analysis. The sample data on each plate is then multiplied by the normalization factor to obtain normalized volume values.

20

B. The Capture/Detection System.

A 96-well dot blot apparatus is typically used to capture the DNA:protein complexes on a nitrocellulose filter as described in Example 10.

25

C. Quantitation of Data.

The autoradiographs of the nitrocellulose filters are analyzed as described in Example 10.

30

D. Analysis of Data.

After densitometry, the data is analyzed using a spreadsheet program, such as "EXCEL." For each plate, the calibrator samples are examined and used to determine the normalization value. Then, for each test

210

oligonucleotide, at each drug concentration and/or each time point, a normalized % score is calculated. The normalized % score (n%) can be described as follows:

$$n\% = (nT/nC) \times 100,$$

5 where (i) nT is the densitometry volume of the test sample multiplied by the normalization factor for the plate from which the sample was obtained, and (ii) nC is the densitometry volume of the control sample multiplied by the normalization factor for the plate
10 from which the sample was obtained. The oligonucleotides are then ranked from 1 to 256 based on their n% scores.

While the invention has been described with
15 reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

211

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Genelabs Technologies, Inc.
- (B) STREET: 505 Penobscot Drive
- (C) CITY: Redwood City
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) POSTAL CODE: 94063

(ii) TITLE OF INVENTION: Sequence-Directed DNA Binding
Molecules, Compositions and Methods

(iii) NUMBER OF SEQUENCES: 641

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Genelabs Technologies, Inc.
- (B) STREET: 505 Penobscot Drive
- (C) CITY: Redwood City
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) ZIP: 94063

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/123,936
- (B) FILING DATE: 17-SEP-1993

(viii) PRIOR APPLICATION DATA:

212

- (A) APPLICATION NUMBER: US 07/996,783
- (B) FILING DATE: 23-DEC-1992

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Fabian, Gary R.
- (B) REGISTRATION NUMBER: 33,875
- (C) REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (415) 324-0880
- (B) TELEFAX: (415) 324-0960

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human ferredoxin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCTCTGCTTG CCAATGTCTT TATAGGTCAC CCGGAAGGCA CG

42

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

213

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human macrophage alpha1-antitrypsin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCTACTGCCT CCACCCGAAG TCTACTTCCT GGGTGGGCAG GAAC

44

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene B for alpha 1-acid
glycoprotein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGTGACCGCC CATAGTTTAT TATAAAGGTG ACTGCACCCT GCAGCC

46

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

214

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for alpha 1
microtubulin-bikunin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATTGGAGCTG TCCTTGGGGC TGTAATTGGC CCCAGCTGAG CAGGGCA

47

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for alpha-2 macroglobulin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTGTTTGAC ACAGAGCAGC ATAAAGCCCA GTTGCTTTGG GAAGT

45

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

215

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ACAA gene for peroxisomal
3-oxoacyl-CoA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTCGGGTTTG GCTACAAAAG GTGGAAAGAC TTCCGGTCTG CATTCTG

48

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ACAA gene for peroxisomal
3-oxoacyl-CoA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAAGGTAGGC GGGGCATTGA GTGGAAAGCT CGGCTGGGCG GTGCCTGT

48

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

216

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human choline acetyltransferase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCAATTGTGA CCCACAGCCT AATAATAACA GTCTTTGCCC TCTTGGCC

48

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human angiotensin I-converting
enzyme gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGCGGGGGTG TGTCGGGTTT TATAACCCGC AGGCGGGCCG CGGCG

45

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

217

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (vi) ORIGINAL SOURCE:
 - (C) INDIVIDUAL ISOLATE: Human gene fragment for the acetylcholine receptor gamma

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGGGTGGGAG TGTAGGCTGT TATATGACAC CCAGAGCCCA TCTCT

45

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

- (vi) ORIGINAL SOURCE:
 - (C) INDIVIDUAL ISOLATE: Human cytokine (Act-2) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTCCTAGGCC TCAGAGTCCC TATAAAGAGA GATTCCCAAC TCAGTA

46

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

218

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human beta-actin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGTGAGTGAG CGGCGCGGGG CCAATCGCGT GCGCCGTTCC GAAAG

45

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human beta-actin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGCGGCCGC GCGGCGCCC TATAAAACCC AGCGGCGCGA CGCGCCA

47

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

219

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cardiac actin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TGCTCCTCAACT GACCCTGTCC ATCAGCGTTC TATAAAGCGG CCCTCCTGGA

50

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for vascular smooth
muscle alpha-actin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAGGAGAGCA GGCCAAGGGC TATATAACCC TTCAGCTTTC AGCTTCC

47

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

220

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human enteric smooth muscle
gamma-actin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AAGATCCGCC TCTGGGGTTT TATATTGCTC TGGTATTCAT GCCA

44

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human arachidonate 12-lipoxygenase
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCGGGGCCGC AGACCGGTCC TTAAAGGTT GGAAGTGGCC CCGAGG

46

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

221

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alcohol dehydrogenase alpha
subunit (ADH1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGTGTTATTC AAGCAAAAAA AATAAATAAA TACCTATGCA ATACACCT 48

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alcohol dehydrogenase beta
subunit gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GATGTTACAC AAGCAAACAA AATAAATATC TGTGCAATAT ATCTGCTT 48

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

222

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-fetoprotein gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TAACAGGCAT TGCCTGAAAA GAGTATAAAA GAATTCAGC ATGATTTTCC 50

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytosolic adenylate kinase
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGCCGCGCG CTGACAGCCT TATAAATAGT CGCCTTTGCC GGCCGCC 47

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

223

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human
alpha-N-acetylgalactosaminidase (AK1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CGGACTTATC AGGTTACCGG ATTCGAGTCA GAAGCGGCGG CAGGTCTGAA

50

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ALAD gene for porphobilinogen
synthase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ATAAAGACCT TTGATCGGAT CTATCATTGT ACCTATCATA GGTCTG

46

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

224

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ALAD gene for porphobilinogen synthase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCCTACCAAG GAGGAAGACT GGATAAAATG GCCTGAGATG GCTGAA

46

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human albumin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAGAGTGACA AAGGCCTGAA TTTGTCAATT AGTAACAATT GTATTCAACA GTAAGGAT 58

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

225

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTGCTCACCA CACACAAGTG TTATAGGAGG AGTCTGGCCC TTGAG

45

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase C gene for fructose
1,6-bisphosphate aldolase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ACCTGCAATA CCCCCTTACC CCAATACCAA GACCAACTGG CATAG

45

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

226

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase C gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGCATAGAGC CAACTGAGAT AAATGCTATT TAAATAAAGT GTATTTAATG AATTCTCCA 60

AGCTTACGGA 70

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CCTCCACACG TCAACGATTC TATTGAAGT TGGGCAGGGG GTGGC 45

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

227

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATTAGAGAAG ATCGGGGACA CATGTGGGGC TGGGCAGGAG CTG

43

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GGGCTGGGCA GGAGCTGCCT TATAACCACC CGGGAACCCC TAGCT

45

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

228

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCGGAGGGCG GAGTGGTGCC TTTAAAAGGC CGGCGCCGCC TTCCGC

46

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TGCGCCGCC CTTCCGAGGC TAAATCGCTT CCTCTCGGAA CGCGC

45

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

229

(C) INDIVIDUAL ISOLATE: Human aldolase B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AAAAAACATG ATGAGAAGTC TATAAAAATT GTGTGCTACC AAAGA

45

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human leukemia inhibitory factor
(LIF) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CTTACAACAC AGGCTCCAGT ATATAAATCA GGCAAATTCC CCATTGAGC

50

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aminopeptidase N gene

230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGGGCTCCTC CCCTTTGGGG ATATAAGCCC GGCCTGGGGC TGCTCC

46

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-amylase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AAATGTGCTT CTTACAGGAA TATAATAGT TTCTGGAAAG GACACTG

47

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human amyloid-beta protein gene

231

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGGAGGCCTG CGGGGTCGGA TGATTCAAGC TCACGGGGAC GAGCAGG

47

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human amyloid beta protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CGGGGACGAG CAGGAGCGCT CTCGACTTTT CTAGAGCCTC AGCGTCCTAG GACT

54

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human amyloid-beta protein

232

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCGGGGTGGG CCGGATCAGC TGACTCGCCT GGCTCTGAGC CCCGCCG

47

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human amyloid-beta protein (APP)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

TCAGCTGACT CGCCTGGCTC TGAGCCCCGC CGCCGCGCTC GGGCTCCGTC

50

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pronatriodilatin precursor
gene

233

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TGCTTGGAGA GCTGGGGGGC TATAAAAAGA GCGGGCACTG GGCAGC

46

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for atrial natriuretic factor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TTGAAGTGGG AGCCTCTTGA GTCAAATCAG TAAGAATGCG GCTCTTGCA

49

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for atrial natriuretic factor

234

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CTGCCGATGA TAAC TT TAAA AGGGCATCTC CTGCTGGCTT CTCAC TTGG

49

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for atrial natriuretic factor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TGCTTGGAGA GCTGGGGGGC TATAAAAAGA GGCGGCACTG GGCAGC

46

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human atrial natriuretic factor

235

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CTTGGAGAGC TGGGGGGCTA TAAAAAGAGG CGGCACTGGG CAGCTGGGAG 50

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human angiotensinogen gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CTCCATCCCC ACCCCTCAGC TATAAATAGG GCCTCGTGAC CCGGCC 46

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human heart/skeletal muscle ATP/ADP
translocator gene

236

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TCGCGAGAGC CCGGCGGGGA TATAAGGGGG AGCTGCGGGC CAGGC

45

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein CIII gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TCTGGACACC CTGCCTCAGG CCCTCATCTC CACTGGTCAG CAGGTGACC

49

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein CIII gene

237

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CTCAGGCCCT CATCTCCACT GGTCAGCAGG TGACCTTTGC CCAGCGCCC

49

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein CIII gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TGCCTGCTGC CCTGGAGATG ATATAAACA GGTCAGAACC CTCCTGCC

48

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein CIII gene

238

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GACACCCTGC CTCAGGCCCT CATCTCCACT GGTCAGCAGG TGACCTTTGC

50

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein CIII gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TGCCTGCTGC CCTGGAGATG ATATAAAACA GGTCAGAACC CTCCTGCC

48

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein AII gene

239

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

ATAATCCCTG CCCCACTGGG CCCATCCATA GTCCCTGTCA CCTGACAGG

49

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein AII gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GGGGTGGGTA AACAGACAGG TATATAGCCC CTCCTCTCC AGCCAG

46

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human fetal gene for apolipoprotein
AI precursor

240

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CTGCAGACAT AAATAGGCCC TGCAAGAGCT GGC

33

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GCCTGGGCTT CCTATAAATG GGGTGC GGCG GCCGGCCGC

39

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apoC-II gene for
preproapolipoprotein C-II

241

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CGGAAGTGGG TCTCAACCAC TATAAATCCT CTCTGTGCCC GTCCGGA

47

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein C-I (VLDL) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TGCCCCGCCC CTCCCCAGCC TGATAAAGGT CCTGCGGGCA GGACAGG

47

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apolipoprotein D gene

242

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ATCAGAGACC TGAAGAAGCT TATAAAATAG CTTGGGAGAG GCCAGTC

47

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human arginase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGTTGTTTAT TCAACCCAAG TATAAATGGA AAAAAAAGAT GCGCC

45

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human argininosuccinate synthetase
gene

243

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CTGCCCCCGG GCCCTGTGCT TATAACCTGG GATGGGCACC CCTGC

45

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human sodium/potassium ATPase alpha
3 subunit (ATP1 A3)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CCCCTCCCGC GGACGCGGGC ATATGAGGAG GCGGAGGCGG CGGC

44

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human (BSF-2/IL6) gene for B cell
stimulatory factor-2

244

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

ATTAGAGTCT CAACCCCCAA TAAATATAGG ACTGGAGATG TCTGAGGC

48

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human C5 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TCTGAATTCT TCAAGTTCAG TTTATTTAAA AGGAGACTAT CCTCAAAAGT G

51

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human carbonic anhydrase II gene

245

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CCTCCCCTTG TCGCCTAGGT CCACCCGAGC CCCCTCCCCC GGGCC

45

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human carbonic anhydrase II gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GCACGAAGTT GGC GGGAGCC TATAAAAGCG GGCCGGCGCG ACCCGC

46

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human calcitonin/alpha-CGRP gene

246

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TTCCCCGACCC ACAGCGGCGG GAATAAGAGC AGTCGCTGGC GCTGG

45

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human calretinin gene, exon 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CAGGCGCAGG CTCCAGAGCG TATATAAGGG CAGCGTGGCG CACAACC

47

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cathepsin G gene

247

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

TTCCTTCCTC TCTCAGGGCC TTAAAGTCTA GGAGGAGGAA GCACA

45

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human carbonic anhydrase VII (CA VII) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CTCCTCCCGC CAGCCGCTGC TTAAGAGGC TGCTCCGCGG TAGCG

45

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cardiac beta myosin heavy chain gene

248

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TCTAGTGACA ACAGCCCTTT CTAAATCCGG CTAGGGACTG GGTGCC

46

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cardiac beta myosin heavy chain gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

TGGGGGTGCC TGCTGCCCCA TATATACAGC CCCTGAGACC AGGTC

45

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human complement C3 gene

249

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

TGGGGGAAAG CAGGAGCCAG ATAAAAAGCC AGCTCCAGCA GGCGCTG

47

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human recognition/surface antigen
(CD4) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CAAGTCCTCA CACAGATACG CCTGTTTGAG AAGCAGCGGG CAAGAAAGAC

50

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human hyaluronate receptor gene
(CD44)

250

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

TAGGTCAGTGG TTTTCAACCT CGAATAAAAA CTGCAGCCAA CTTCCGAGGC

50

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cystic fibrosis transmembrane conductance reg. gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

AATGACATCA CAGCAGGTCA GAGAAAAAGG GTTGAGCGGC AGGCACCCAG

50

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cholesterol 7-alpha-hydroxylase (CYP7) gene

251

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

ATGGATCTGG ATACTATGTA TATAAAAAGC CTAGCTTGAG TCTCTT

46

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human choline acetyltransferase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

AGCAATTGTG ACCCACAGCC TAATAATAAC AGTCTTTGCC CTCTTGCC

49

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human mast cell chymase gene

252

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

CTCTCTTGCC TTCTAGGAGT TATAAAACCC AAGACTGGAA AGGAAA

46

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human heart chymase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CCTCTCTTGC CTTCTGGGAG TTATAAAACC CAAGACTGGA AGGAAAA

47

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human creatine kinase B gene

253

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

GGCCAATGGA ATGAATGGGC TATAAATAGC CGCCAATGGG CGGCCCCG

48

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human C-type natriuretic peptide gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ACATCAGCGG CAGGTTGGAT TATAAAGGCG CGAGCAGAGT CACGGG

46

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transmembrane protein (CD59) gene

254

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GCTCCGCGCG GGGGTGGAGG GAGAGGAGGA GGTTCCTGCC GAGGT

45

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transmembrane protein (CD59)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

GAGGGCAAGG GCATCCTGAG GGGCGGGGCC GGGGGCGGAG CCTTGC

46

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transmembrane protein (CD59)
gene

255

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

ATCCTGAGGG GCGGGGCCGG GGGCGGAGCC TTGCGGGCTG GAGCGA

46

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transmembrane protein (CD59)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

TGAGGGGCGG GGCCGGGGGC GGAGCCTTGC GGGCTGGAGC GAAAGAATGC

50

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human myeloid specific CD11b gene

256

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GCCCTCTTCC TTTGAATCTC TGATAGACTT CTGCCTCCTA CTTCTC

46

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cholesteryl ester transferase
protein (CETP) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

GTGGGGGCTG GCGGGACATA CATATACGGG CTCCAGGCTG AACGGC

46

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cystic fibrosis transmembrane
conductance regulator

257

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TGGGTGGGGG GAATTGGAAG CAAATGACAT CACAGCAGGT CAGAG

45

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cystic fibrosis transmembrane
conductance regulator

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GTGGGGGGGAA TTGGAAGCAA ATGACATCAC AGCAGGTCAG AGAAAAA

47

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human coseg gene for
vasopressin-neurophysin precursor

258

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

CACGGGAACA CCTGCGGACA TAAATAGGCA GCCAGCAGAG GCAGCA

46

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human creatine kinase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

TTCAGAGAAA GGGCAGGTGC TATAAAGGGC CCAGCGCCAC GGGCCT

46

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-B-crystallin gene

259

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

AGAAGCTTCA CAAGACTGCA TATATAAGGG GCTGGCTGTA GCTGCAG

47

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human C3 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

AGTGGGGGAA ACCAGAGCCA GATAAAAAGC CAGCTCCAGC AGGCGCTGCT

50

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human colony stimulating factor
CSF-1 gene

260

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

GCCTGGCCAG GGTGATTTC CATAAACCAC ATGCCCCCA GTCCTC

46

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytotoxic serine proteinase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GTTACTCAGC AGCAGGGGTG TAAATGTGAC AGTGCCATGT CAAC

44

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human CST3 gene for cystatin C

261

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GGCGGCGAAG GCCGGAAGGG ATAAAACCGC AGTCGCCGGC CTCGCG

46

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human CST4 gene for Cystatin D

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

TTGGGGGACA CCCAAGTAGG ATAAATGCAC AGCTAGCTTC TGGCC

45

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human CYP2C8 gene for cytochrome
P-450

262

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

ACTAAATTAG CAGGGAGTGT TATAAAACT TTGGAGTGCA AGCTC

45

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cholesterol desmolase
cytochrome gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

AGCAGGAGGA AGGACGTGAA CATTTTATCA GCTTCTGGTA TGGCC

45

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cholesterol desmolase
cytochrome P-450 gene

263

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

TATGGCCTTG AGCTGGTAGT TATAATCTTG GCCCTGGTGG CCCAGG

46

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human steroid 11-beta-hydroxylase
(CYP11B1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

GAAGGCAAGG CACCAGGCAA GATAAAAGGA TTGCAGCTGA ACAAGGT

47

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human CYPXI gene for steroid
18-hydroxylase (P-450 C18)

264

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

CAGAGCAGGT TCCTGGGTGA GATAAAAGGA TTGGGGCTGA ACAGGGT

47

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human CYPXIX aromatase P-450 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

TGGACAATAA ATGAAATCTC CATAAAAGGC CCAAAGGACA GGGTTC

46

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human decay-accelerating factor
(DAF) gene

265

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

AGCCCAGACC CCGCCCAAAG CACTCATTTA ACTGGTATTG CGGAGC

46

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human dopamine beta-hydroxylase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

ACGTCCATGT GTCATTAGTG CCAATTAGAG GAGGGCAGCA GGCTG

45

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human dopamine beta-hydroxylase gene

266

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

ACCCCATTC A GGACCAGGGC ATAAATGGCC AGGTGGGACC AGAGAG

46

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human desmin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

GGGGCTGATG TCAGGAGGGA TACAAATAGT GCCGACGGCT GGGGGC

46

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytokeratin 8 (CK8) gene

267

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

CCCGGGGCTG GGATCTCTTT TATAAAAGGC CATTCTGAG AGCTC

45

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human DNA polymerase alpha gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

GCCTCCCGAG CCGCTGATTG GCTTTCAGGC TGGCGCCTGT CTCGGCCCCC

50

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human dopamine D1A receptor gene

268

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

GCTGTGCCCC CCGGGAACCC CGCCGGCCTG TCGCCTTGCT GGTGCCAGCT 50

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human eosinophil cationic protein (ECP) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

AGACCCACCA AGGGAAGCTT TATTAAACA GTTCCAAGTA GGGGAGA 47

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human HER2 gene

269

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

GAGGAGGAGG GCTGCTTGAG GAAGTATAAG AATGAAGTTG TGAAGCTGAG

50

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human elastin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GTGTCTCGCT GTGATAGATC AATAAATATT TTATTTTTTG TCCTGG

46

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human endothelial leukocyte adhesion molecule I (ELAM-1)

270

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

ATTACAGGA AGCAATCCCT CCTATAAAG GCCCTCAGCC GAAGTAGTG

49

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human eosinophil major basic protein gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

GGAAGTTCCT CCAAGGCCTC TATATAAGAA GTCTTTGTGA GAGGAAG

47

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human preproenkephalin B gene

271

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

CTCTAGGAAA GTTTCCTCAGC TCTCAAACCT CTGTTTCTC ATCTGCAAG

49

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human preproenkephalin B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

TTTCATCTG CAAGATGGGG ATAATATTAA CCAACTGGCT AGGTCATGAG

50

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human ENO3 gene for muscle-specific enolase

272

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

GGGGACCGAG TGGCTCAGGG ATAAATGCGC ACCTGAGAGG GGGTGA

46

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human eosinophil derived neurotoxin
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

CAACCCACCA AGGGATGCTT TATTTAAACA GTTCCAAGTA GGGGAGA

47

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human erythropoietin receptor

273

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

TACCCAGGCT GAGTGCTGGC CCCGCCCCCT CGGGGATCTG CCACTT

46

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human c-erb B2/neu protein gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

AGGAGGGCTG CTTGAGGAAG TATAAGAATG AAGTTGTGAA GCTGA

45

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ERCC2 gene

274

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

CCGATTGGCT CTGCCCTAGC GGATTGACGG GCAGGTTAGC CAATGGTCT

49

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ERCC2 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

CAGGTTAGCC AATGGTCTCG TAATATAGGT GGAGCGAGCC CTCGAGG

47

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human erythropoietin gene

275

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

GGTCACCCGG CGCGCCCCAG GTCGCTGAGG GACCCCGGCC AGGCGCGGAG

50

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human oestrogen receptor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

ATATGAGCTC GGGAGACCAG TACTTAAAGT TGGAGGCCCG GGAGCCCA

48

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human elastase I gene

276

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

AGCTTTGCTG CTAAGAGGAG TATAAGAGG GCTTGGTCCA AGCAAG

46

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

GGCCCCGTGA TCAGCTCCAG CCATTTCAG TCCTGGCTAT CCA

44

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene

277

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

TGGCTATCCC AGGAGCTTAC ATAAAGGGAC AATTGGAGCC TGAGA

45

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human lymphocyte IgE receptor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

TTAACATCTC TAGTTCTCAC CCAATTCTCT TACCTGAGAA ATGGA

45

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human lymphocyte IgE receptor gene

278

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

GTTATCCGGG TGGCAAGCCC ATATTTAGGT CTATGAAAAT AGAAGCT

47

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human lymphocyte IgE receptor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

AGCCCATATT TAGGTCTATG AAAATAGAAG CTGTCAGTGG CTCTAC

46

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human apoferritin H gene

279

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

GGGCCTGACG CCGACGCGGC TATAAGAGAC CACAAGCGAC CCGCA

45

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human fibrinogen beta gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

TATTAAGTAA GGAAAGGTAA CCATTCTGA AGTCATTCCT AGCAGA

46

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human fibrinogen beta gene

280

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

ATTCCTAGCA GAGGACTCAG ATATATATAG GATTGAAGAT CTCTCAGTT

49

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human factor IX gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CCAGAAGTAA ATACAGCTCA GCTTGTACTT TGGTACAACCT AATCGACCTT

50

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human FK506 binding proteins 12A,
12B, and 12C (FKBP12)

281

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

GAGCCGTGGA ACCGCCGCCA GGTCGCTGTT GGTCCACGCC GCCCGTCGCG 50

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human 5-lipoxygenase activating protein (FLAP) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

TTGTGCCGGG GATCTTCAGA AATTGTAATG ATGAAAGAGT GCAAGCTCTC 50

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human fos proto-oncogene (c-fos)

282

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

ATTCATAAAA CGCTTGTTAT AAAAGCAGTG GCTGCGGCGC CTCGTACTCC

50

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human GOS2 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GGCGTGTCTC AGAGAAAAGA TATAAGCGGC CCCC GGACGC TAAAG

45

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human granulocyte colony-stimulating
factor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

CAGGCCTCCA TGGGGTTATG TATAAAGGGC CCCCTAGAGC TGGGCC

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human EGR2 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

CGGGTATTGA AGACCTGCCC ATAAATACTT AGAGCAACAC TTTCCGTC

48

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human growth hormone (hGH) gene

284

DESCRIPTION: SEQ ID NO:144:

TATAAAAAGG GCCCACAAGA GACCAG

46

SEQ ID NO:145:

CHARACTERISTICS:

(i) LENGTH: 50 base pairs

(b) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gastric inhibitory polypeptide
(GIP) mRNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

TAATCAGCAG GTCTATGCCT AATATAAAGG AGCTGGGGCA TGATTCTTC

50

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human GLA gene for
alpha-D-galactosidase A

285

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GAAACAATAA CGTCATTATT TAATAAGTCA TCGGTGATTG GTCCGC

46

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human glucagon gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

TTTACAGATG AGAAATTTAT ATTGTCAGCG TAATATCTGT GAGG

44

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human glucagon gene

286

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

GGCTAAACAG AGCTGGAGAG TATATAAAG CAGTGCGCCT TGGTGCA

47

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human granulocyte-macrophage colony stimulating factor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

CATTAAATCAT TTCCTCTGTG TATTTAAGAG CTCTTTTGCC AGTGAGC

47

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human glucocorticoid receptor gene

287

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

TGGGCAATGG GAGACTTTCT TAAATAGGGC TCTCCCCCA CCCATG

46

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human growth hormone releasing factor (GRF) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

AACGCTTAGG AAAATGAAGA GATAAATGAT GGGAACGCCA GCGGCTGCC

50

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human GST pi gene for glutathione S-transferase pi

288

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GAGCGGGGCG GGACCACCCT TATAAGGCTC GGAGGCCGCG AGGC

44

(2) INFORMATION FOR SEQ ID NO:153:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: human glycoporphin C (GPC) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

CAGAAGTGGG CGGGTGTGTG TTAAAAAAA AAAAAAGGGG TGGAAAC

47

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone (H10) gene

289

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

CGCGGTCCGC CCGCCGCCGC TAAATACCCG GATGCGCCGC CCAAGC

46

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for H1 RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

GTCTTTGGAT TTGGGAATCT TATAAGTTCT GTATGAGACC ACTC

44

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human H1 histone gene FNC16

290

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

GGCGGTGGAT TGGACGCTCC ACCAATCACA GGCAGCGCC GGCTTA 46

(2) INFORMATION FOR SEQ ID NO:157:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone gene FNC16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

ACCAATCACA GGCAGCGCC GGCTTATATA AGCCCGGGCC CGAGCATAGC AGCA 54

(2) INFORMATION FOR SEQ ID NO:158:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human H2B.2 and H2A.1 genes for
Histone H2A and H2B

291

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

TTTTCGCGCC CAGCAGCTGC TATAAATGC GCGTCCCTGT AGGTTCC

47

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human H4/a gene for H4 histone

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

GGGGGCAGGG GTAACGTAGA TATATAAAGA TCGGTTTCCT ATTCTCTC

48

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human H4/b gene for H4 histone

292

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

CTGCAAGTAT AGTGTGTGTG TATATATATA TATATACCTA GCAGTATTTA TTAAAT 56

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human androgen receptor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

GGTGGGGGCG GGACCCGACT CGCAAAGTGT TGCATTTGCT CTCCACCTCC 50

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human chorionic gonadotropin (hCG)
beta subunit

293

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

GCCCTCTCTC ATTGGGCAGA AGCTAAGTCC GAAGCCGCGC CCCTCCTGG

49

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human islet amyloid polypeptide (hIAPP) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

GCTGAGAAAG GTGTGAGGGG TATATAAGAG CTGGATTACT AGTTAGCAAA

50

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human H4 histone gene

294

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

CTTCCCGCCG GCGCGCTTTC GGTTTCAAT CTGGTCCGAT ATCTCTGTAT AT 52

(2) INFORMATION FOR SEQ ID NO:165:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human H4 histone gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

AATCTGGTCC GATATCTCTG TATATTACGG GGAAGACGGT GACGCTC 47

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H2a gene

295

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

TCCTCTTTTC TTGGCGAACT CAACTGGTAT GAATTCCTCA

40

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H2a gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

CACAGCCTAC CTCCAGTCAG TATAAATACT TCTCTGCCTT GCGTTC

46

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H2b gene

296

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

TATTTCATA AGCGATTCTA TATAAAGCG CCTGTCATA CCCTGCT

47

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H3 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

ATTTTGAAT TTTCTTGGGT CCAATAGTTG GTGGTCTGAC TCTAT

45

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H3 gene

297

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

CAATAGTTGG TGGTCTGACT CTATAAAGA AGAGTAGCTC TTCCTT

47

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human HLA-A1 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

AGTGTCGTCG CGGTCGCTGT TCTAAAGTCC GCACGCACCC ACCGG

45

(2) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human HLA-B27 antigen gene

298

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

AGTGTGCGCCG GGGTCCCACT TCTAAAGTCC CCACGCACCC ACCCGG

46

(2) INFORMATION FOR SEQ ID NO:173:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human HLA-Bw57 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

AGCGTCGCCC GGGTCCCACT TCTAAAGTCC CCACGCACCC ACCCG

45

(2) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human HLA-F gene for human leukocyte antigen F

299

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

TGTCGCCGCA GTTCCCAGGT TCTAAAGTCC CACGCACCCC GCGGGA

46

(2) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for histocompatibility antigen HLA-A3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

AGTGTCGTCG CGGTCGCTGT TCTAAAGCCC GCACGCACCC ACCGGG

46

(2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for class I histocompatibility antigen HLA-CW3

300

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

CATTGGGTGT CGGACCTCTA GAAGGCCGGT CAGCGTCTCC GC

42

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human HMG-17 gene for non-histone
chromosomal protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

CGGTCCGGGG CTCCAGCGC TATAAAACT TTATAAACCC CCCGGA

46

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human HOX3D gene for homeoprotein
HOX3D

301

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

AAGAAAGAGA TATCTCCACC TATAAATTGT CCACTTTGGA GAACAA

46

(2) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human 71Kd heat shock cognate protein (hsc70)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

TGGAAGGTTT TAAGATAGGG TATAAGAGGC AGGCTGGCGG GCGGA

45

(2) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human heat shock protein (hsp 70) gene

302

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

AAGGCGGGTC TCCGTGACGA CTTATAAAAG CCCAGGGGCA AGCGGTCCGG

50

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hsp70B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

CTTCGGTCTC ACGGACCGAT CCGCCCGAAC CTTCTCCCGG GGTCAG

46

(2) INFORMATION FOR SEQ ID NO:182:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hsp70B gene

303

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

CCGCCCCGGCT GACTCAGCCC GGGCGGGCGG GCGGGAGGCT CTCGAC

46

(2) INFORMATION FOR SEQ ID NO:183:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hsp70B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CCGGCTGACT CAGCCCCGGC GGGCGGGCGG GAGGCTCTCG ACTGGG

46

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hsp70B gene

304

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

CTGACTCAGC CCGGGCGGGC GGGCGGGAGG CTCTCGACTG GGC GGG

46

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hsp70B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

GGGCGGGCGG GAGGCTCTCG ACTGGGCGGG AAGGTGCGGG AAGGT

45

(2) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hsp70B gene

305

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CGGCGGGGTC GGGGAGGTGC AAAAGGATGA AAAGCCCGTG GACGGAGCTG AGC 53

(2) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human IAPP gene for islet amyloid polypeptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

GCTGAGAAAG GTGTGAGGGG TATATAAGAG CTGGATTACT AGTTAGC 47

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human intercellular adhesion molecule 1 (ICAM-1) gene

306

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

AGGTTTCCGG GAAAGCAGCA CCGCCCCTTG GCCCCCAGGT GGCTAG

46

(2) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human intercellular adhesion molecule 1 (ICAM-1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

GGCCCCCAGG TGGCTAGCGG TATAAAGGAT CACGCGCCCC AGTCGA

46

(2) INFORMATION FOR SEQ ID NO:190:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon-inducible gene IFI-54K

307

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

AAAGGAACCA GAGGCCACTG TATATATAGG TCTCTTCAGC ATTTATTG

48

(2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
IFN-alpha 14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

ATGGAAGCTA GTATGTCCT TATTTAAGAC CTATGCACAG AGCAAGGT

48

(2) INFORMATION FOR SEQ ID NO:192:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
IFN-alpha 16

308

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

GAAATTAGTA TGTTCACTAT TTAAGAACTA TGCACAGAGC AAAGT

45

(2) INFORMATION FOR SEQ ID NO:193:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
IFN-alpha 5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

ATGGAAACTC GTATGTGACC TTTTAAAGAT CTGTGCACAA AACCAAGGT

48

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
IFN-alpha 6

309

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

ATGGAACTA GTATGTTCCC TATTTAAGAC CTACACATAA AGCAAGGT

48

(2) INFORMATION FOR SEQ ID NO:195:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
IFN-alpha 7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

ATGGAAATTA GTATGTTTAC TATTTAAGAC CTATGCACAG AGCAAAGT

48

(2) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human immune interferon (INF-gamma)
gene

310

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

TCCTCAGGAG ACTTCAATTA GGTATAAATA CCAGCAGCCA GAGGAGGTGC

50

(2) INFORMATION FOR SEQ ID NO:197:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human alpha/beta-interferon
(IFN)-inducible 6-16 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

GGGAGGATCC ACAAGTGATG ATAAAAAGCC AGCCTTCAGC CGGAG

45

(2) INFORMATION FOR SEQ ID NO:198:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human insulin like growth factor II
(IGF-2)

311

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

CTGGGAGGAG TCGGCTCACA CATAAAGCT GAGGCACTGA CCAGCCT

47

(2) INFORMATION FOR SEQ ID NO:199:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human insulin-like growth factor binding protein gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

GTGGCGCGGC CTGTGCCCTT TATAAGGTGC GCGCTGTGTC CAGCG

45

(2) INFORMATION FOR SEQ ID NO:200:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human germline leader peptide and variable region of 1154

312

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

CAACCTCCTG CACTGAAGCC TTATTAATAG GCTGGCCACA CTCATGC 48

(2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human germline for leader peptide & variable region of 2908

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

CAACCTCCTG CCCTGAAGAC TTATTAATAG GCTGGACACA CTCATGC 48

(2) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human rearranged kappa immunoglobulin subgroup V

313

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

CCACGACCAG GTGTTTGGAT TTTATAAACG GGCCGTTTGC ATTGTGAA

48

(2) INFORMATION FOR SEQ ID NO:203:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human rearranged kappa
immunoglobulin gene subgroup V

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

CGCCCTGCAG TCCAGAGCCC ATATCAATGC CTGGGTCAGA GCTCTGGA

48

(2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human germline fragment for
immunoglobulin kappa light chain

314

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

TGCCCTACCT TCCAGAGCCC ATATCAATGC CTGTGTCAGA GCCCTGGG

48

(2) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human germline immunoglobulin kappa
light chain V-segment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

ACTTCCCTTG TGGGTCTGAG ATAAAAGCTC AGCTCTAACC CTTACC

46

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin-2 (IL-2) gene

315

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

TATTTTCCA GAATTAACAG TATAAATTGC ATCTCTGTT CAAGAG

46

(2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for interleukin 1 alpha
(IL-1 alpha)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

CCACGCCTAC TTAAGACAAT TACAAAAGGC GAAGAAGACT GACTCAG

47

(2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for prointerleukin 1 beta

316

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

TTGATTGTGA AATCAGGTAT TCAACAGAGA AATTTCTCAG CCTCCTAC

48

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for prointerleukin 1 beta

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

CTACTTCTGC TTTTGAAAGC TATAAAAACA GCGAGGGAGA AACTGGC

47

(2) INFORMATION FOR SEQ ID NO:210:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: human interleukin 2 receptor gene

317

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

AGAAACGATT CATAAATGAA GTTCAATCCT TCTCATCACC CCAGCCCA

48

(2) INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 2 receptor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

TTTGAAAAAT TACCGCAAAC TATATTGTCA TCAAAAAAAA AAAAAA

46

(2) INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 4 gene

318

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

ATCTGGTGTA ACGAAAATTT CCAATGTAAA CTCATTTTCC CTCGG

45

(2) INFORMATION FOR SEQ ID NO:213:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 4 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

GGTTTCAGCA ATTTTAAATC TATATATAGA GATATCTTTG TCAGCATT

48

(2) INFORMATION FOR SEQ ID NO:214:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 5 (IL-5) gene

319

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

CATTTCCTCA AAGACAGACA ATAAATTGAC TGGGGACGCA GTCTTGACT

50

(2) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 7 (IL-7) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

TTGCTTTGAT TCAGGCCAGC TGGTTTTTCT GCGGTGATTC GGAAATTCGC

50

(2) INFORMATION FOR SEQ ID NO:216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 9 gene (IL-9)

320

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

TTCCGTGTTT GAGAGGGAGC TTAAATACC ACTCGATTG AAGGTGTC

48

(2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human int-1 mammary oncogene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

ACTTCAGCCA GCGCCGCAAC TATAAGAGGC GGTGCCGCC CCGT

45

(2) INFORMATION FOR SEQ ID NO:218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human jun-B gene

321

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

TCCGTGGCTG ACTAGCGCGG TATAAAGGCG TGTGGCTCAG GCTGAG

46

(2) INFORMATION FOR SEQ ID NO:219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human DNA for 65 kD keratin type II

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

GCCCAACAAC CTCCTCAAAT GTATATAAAG GGATTTTAT TGCACA

46

(2) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ultra high-sulphur keratin
protein gene

322

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

TGGTGTGTTC CTATGTGGGA TATAAAGAGC CGGGGCTCAG GGGGCT

46

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-lactalbumin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

CCTGAGGCTT TCTGCATGAA TATAAATAAA TGAACTGAG TGATGCT

47

(2) INFORMATION FOR SEQ ID NO:222:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human LAG-1 gene

323

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

GTCCTAGGCC TCAGAGTCCC TATAAGAGAG ATTCCCAACT CAGTA

45

(2) INFORMATION FOR SEQ ID NO:223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human lecithin-cholesterol
acyltransferase (LCAT) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

CTGAGGCTGT GCCCCTTTCC GGCAATCTCT GGCCACAACC CCCACTGG

48

(2) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human lecthin-cholesterol
acyltransferase (LCAT) gene

324

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

CCCCTCCCAC TCCCACACCA GATAAGGACA GCCCAGTGCC GCTTT

45

(2) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human lymphocyte-specific protein
kinase (lck) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

GGGAGCAGAT CTTGGGGGAG CCCCTTCAGC CCCCTCTTCC ATTCCTCAG

50

(2) INFORMATION FOR SEQ ID NO:226:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human leukocyte fuction-associated
antigen-1 (LFA-1) gene

325

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

GGGTATCTCA CTGTGGTTTG ATTTGCATTT CTCTAATGAC TAATAGTG

48

(2) INFORMATION FOR SEQ ID NO:227:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human leukocyte function-associated
antigen-1 (LFA-1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

ATGTCTCTAA CTTGCTTACA CTCCTCCCT GAACCCTGCG GTTTCA

46

(2) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human leukocyte function-associated
antigen-1 (LFA-1) gene

326

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

TCCTGCAGGC ACACCTCCCT CCCC GCCTGC CAGTGT CACC AGCCTGTT

48

(2) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human leukocyte function-associated
antigen-1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

CTGTTGCCTC TGTGAGAAAG TACCACTGTA AGAGGCCAAA GGGCATGATC

50

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: human lipoprotein lipase (LPL) gene

327

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

TATTGCATA TTTCCAGTCA CATAAGCAGC CTTGGCGTGA AAACAGT

47

(2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human leukocyte adhesion molecule-1
(LAM-1)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

TGGGTTAGAG AAATGAAAGA AAGCAAGGCT TTCTGTTGAC ATTCAGTGCA

50

(2) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human lysozyme gene

328

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

AGAAGGAAGT TAAAGATGT TAAATACTGG GGCCAGCTCA CCCTGG

46

(2) INFORMATION FOR SEQ ID NO:233:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human mannose binding protein 1 (MBP1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

AGGCATGGGT CATCTATTTT TATATAGCCT GCACCCAGAT TGTAGG

46

(2) INFORMATION FOR SEQ ID NO:234:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human mast cell carboxypeptidase A (MC-CPA) gene

329

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

CATCAAGATA AGGGCTGAGG CATAAACTG CCAGAGGGTC TCAAGG

46

(2) INFORMATION FOR SEQ ID NO:235:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human P-glycoprotein (MDR1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

CTTTGCCACA GGAAGCCTGA GCTCATTCCA GTAGCGGCTC TTCCA

45

(2) INFORMATION FOR SEQ ID NO:236:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human bone marrow serine protease
gene (medullasin)

330

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

ACGGCCTCCC AGCACAGGGC TATAAGAGGA GCCGGGCGGG CACGG

45

(2) INFORMATION FOR SEQ ID NO:237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human microsomal epoxide hydrolase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

TTGCTGTGCA GAGTCCAGGG GAGATAACCA CGCTGTGCAC ACATGAG

47

(2) INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human metallothionein-Ie gene

331

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

GCAGCCAGTT GCAGGGCTCC ATTCTGCTTT CCAACTGCCT GACTGCTTGT TC

52

(2) INFORMATION FOR SEQ ID NO:239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human myoglobin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

TTGTCAAGCA TCCCAGAAGG TATAAAAACG CCCTTGGGAC CAGGCA

46

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human melanoma growth stimulatory
activity (MGSA) gene

332

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

GCTTTCAGC CCCAACCATG CATAAAAGGG GTTCGCGGAT CTCGGAG

47

(2) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human alpha-MHC gene for myosin heavy chain (N-terminus)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

AGAGGGTGGG GGAAACGGGA TATAAAGGAA CTGGAGCTTT GAGGAG

46

(2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human class II invariant gamma-chain gene

333

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

GATTCTCTC CAGCACCGAC TTTAAGAGGC GAGCCGGGGG GTCAG

45

(2) INFORMATION FOR SEQ ID NO:243:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human motilin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

CCCAGGGTTG GGAGGTATAT AAGAACCCGT CAGATCAGCC G

41

(2) INFORMATION FOR SEQ ID NO:244:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human myeloperoxidase gene

334

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

CCACCCCCAG CTTAGAGGAC ATAAAGCGC AGATTGAGCT AAGAGGAGCT

50

(2) INFORMATION FOR SEQ ID NO:245:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human mitochondrial RNA-processing
endoribonuclease RNA gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

AAACACAATT TCTTAGGGC TATAAATAC TACTCTGTGA AGCTGAGGA

49

(2) INFORMATION FOR SEQ ID NO:246:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human myc-oncogene

335

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

GAGGGAGGGA TCGCGCTGAG TATAAAAGCC GGTTCGGG GCTTTAT

47

(2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human Na,K-ATPase beta subunit
(ATP1B) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

GCACGGCCGC CGGGCGCGG TATATAGTAA AGGTAGGGCG GCGCA

46

(2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human neuromedin K receptor

336

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

GAAGCGTGGG ACCCCATGAG TATAAAGAGA GCCTGTAGCG CAGG

44

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for heavy neurofilament subunit (NF-H) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

TTGGACCCGG CCGCGCGCGC TATAAAAGGG CCGCGCCCT GGTCT

46

(2) INFORMATION FOR SEQ ID NO:250:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human nuclear factor NF-IL6 gene

337

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

CGGTTGCTAC GGGCCGCCCT TATAAATAAC CGGGCTCAGG AGAAACT

47

(2) INFORMATION FOR SEQ ID NO:251:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human neurofilament subunit NF-L
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

TGCGTCAGGA CCTCCCGGCG TATAAATAGG GGTGGCAGAA CGGCGC

46

(2) INFORMATION FOR SEQ ID NO:252:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human neurokinin-2 receptor (NK-2)
gene

338

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

TCTCTTCAGC GAAGGGGTTG ATTTATAAGG GTGTTTTCTG CTCTGACA

48

(2) INFORMATION FOR SEQ ID NO:253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human n-myc gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

GGGTGTGTCA GATTTTTCAG TTAATAATAT CCCCCGAGCT TCAAAGCGC

49

(2) INFORMATION FOR SEQ ID NO:254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ornithine decarboxylase (ODC1)
gene

339

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

CCATGGCGAC CCGCCGGTGC TATAAGTAGG GAGCGGCGTG CCGTGG

46

(2) INFORMATION FOR SEQ ID NO:255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human ornithine transcarbamylase
(OTC) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

ATACACAGCG GTGGAGCTTG GCATAAAGTT CAAATCTCC TACACC

46

(2) INFORMATION FOR SEQ ID NO:256:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human prepro-oxytocin-neurophysin I
gene

340

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

CTCCACCGAC GCAATGCCCA GGCATAAAAA GGCCAGGCCG ACAGACCGCC

50

(2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytochrome P450scc gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

TATGGCCTTG AGCTGGTAGT TATAATCTTG GCCCTGGTGG CCCAG

45

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human p53 gene for transmembrane
related p53

341

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

CCCCTCCCAT GTGCTCAAGA CTGGCGCTAA AAGTTTTGAG CTTCTCAAAA

50

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human Alzheimer's disease amyloid A4 precursor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

GGGAGGCCTG CGGGGTCGGA TGATTCAAGC TCACGGGGAC GAGCAGG

47

(2) INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human Alzheimer's disease amyloid A4 precursor gene

342

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

CGGGGACGAG CAGGAGCGCT CTCGACTTTT CTAGAGCCTC AGCGTCCTAG GACT

54

(2) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human Alzheimer's disease amyloid A4 precursor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

GCGGGGTGGG CCGGATCAGC TGA CTGCGCT GGCTCTGAGC CCCGCCGC

48

(2) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human Alzheimer's disease amyloid A4 precursor gene

343

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

CCGCCGCCGC GCTCGGGCTC CGTCAGTTTC CTCGCCAGCG GTAGGCGAG

49

(2) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for plasminogen activator inhibitor 1 (PAI-1)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

TATTTCTGCG CCACATCTGG TATAAAGGA GGCACTGGCC CACAGAG

47

(2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human platelet-derived growth factor A-chain (PDGF) gene

344

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

AGGGGCGCGG CGGCGGCGGC TATAACCCTC TCCCCGCCGC CGGCC

45

(2) INFORMATION FOR SEQ ID NO:265:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human PGP9.5 gene for
neuron-specific ubiquitin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

ACAGTGCCTC TGGCCGGCGC TTTATAGCTG CAGCCTGGCG CTCCGC

46

(2) INFORMATION FOR SEQ ID NO:266:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human plasminogen gene

345

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

CTCCACCGAC GCAATGCCCA GGCATAAAAA GGCCAGGCCG AGAGACCGCC

50

(2) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human phenylethanolamine N-methylase (PNMT) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

TCGGGGCGGG GGTGGGGCGG TAGAAAAAAG GGCCGCGAGG CGAGCGGGG

49

(2) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human opiomelanocortin gene

346

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

CTCCCCGTGT GCAGACGGTG ATATTACCG CCAAATGCGA ACCAGGC

47

(2) INFORMATION FOR SEQ ID NO:269:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene PRB3L for proline-rich protein G1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

GCCACTGTTC TGCTCCTCTT TATAAAGGGA GCTGCCATGG TTCTCC

46

(2) INFORMATION FOR SEQ ID NO:270:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human PRB4 gene for proline-rich protein Po

347

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

CATTGTTTTG CTCCTCTTTA TAAAGGGAGT TGCCACGTTC CTCC

44

(2) INFORMATION FOR SEQ ID NO:271:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human prolactin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

AGGCTTTGAT ATCAAAGGTT TATAAAGCCA ATATCTGGGA AAGAGA

46

(2) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human prothymosin-alpha gene

348

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

CCGAGCGCCG CCCACTAATC TATATTAAAG CTCTGGCGC CGCGTG

46

(2) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human protamine 2 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

TCATAGTGGG CGTCCCCCTT TATATACAAG CTCCCGGGGA GCCTTG

46

(2) INFORMATION FOR SEQ ID NO:274:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human SPR2-1 gene for small proline
rich protein

349

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

CTGGGTGGGG TAGCAGGCTC TATAAAGAGA TCCTCTGCTG CACGAC

46

(2) INFORMATION FOR SEQ ID NO:275:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human estrogen-responsive gene pS2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

TAAGCAAACA GAGCCTGCCC TATAAAATCC GGGGCTCGGG CGGCCTC

47

(2) INFORMATION FOR SEQ ID NO:276:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pulmonary surfactant
apoprotein (PSAP) gene

350

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

AGCCTGGCAG CCCCCACATC TATAAATGCT GCGTCTACCT TACCCT

46

(2) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for prostatic secretory
protein PSP-94

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

TGCGTGGTTG CCCTCTCCAG TATAAPAGTT TGATGCAGCT TTTCC

45

(2) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human parathyroid hormone-related
peptide (PTHrP) gene

351

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

GAGGTAGACA GACAGCTATG TATATATATG TGGGTTTCGC TACAAGTGG

49

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for purine nucleoside phosphorylase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

CTGGGGACTC CAGGGCAAGG GATATAAGCC AGAGCCTAGA CCAAGT

46

(2) INFORMATION FOR SEQ ID NO:280:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human rDNA

352

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

ATTTTGGGCC GCCGGGTTAT

20

(2) INFORMATION FOR SEQ ID NO:281:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human regenerating protein (reg)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

GTTCTATCT CAGATCCTGA TATAAGCTC CTACAGCTAC CTGGCC

46

(2) INFORMATION FOR SEQ ID NO:282:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human renin gene

353

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

ATCACCCCAT GCATGGAGTG TATAAAGGG GAAGGGCTAA GGGAGCC

47

(2) INFORMATION FOR SEQ ID NO:283:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene fragment for retinol
binding protein (RBP)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

CGACCCCCTC CCCCCGGCGC TATAAGCAG CGGGGCGGCC GCGGCG

46

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human serum amyloid A (GSAAl) gene

354

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

CACCCCGCTA ATTTAAAAAA TATATATACA GATATATAGT GGAGATGG

48

(2) INFORMATION FOR SEQ ID NO:285:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human SAA1 beta gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

AACCAGCAGG GAAGGCTCAG TATAAATAGC AGCCACCGCT CCCTGGC

47

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene fragment for HLA class II
SB 4-beta chain

355

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

CTACTTGGGT TCATGGTCTC TAATATTTC AAGAGGAGCT CCCTTTAG

48

(2) INFORMATION FOR SEQ ID NO:287:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human c-sis proto-oncogene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

TCGCACTCTC CCTTCTCCTT TATAAAGGCC GGAACAGCTG AAAGGG

46

(2) INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human SLPI gene for secretory
leukocyte protease inhibitor

356

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

CACACCCACT GGTGAAAGAA TAAATAGTGA GGTTGGCAT TGGCCA

46

(2) INFORMATION FOR SEQ ID NO:289:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human superoxide dismutase (SOD-1)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

CGAGGCGCGG AGGTCTGGCC TATAAAGTAG TCGCGGAGAC GGGGTG

46

(2) INFORMATION FOR SEQ ID NO:290:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ornithine decarboxylase gene

357

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

TCCATGGCGA CCCGCCGGTG CTATAAGTAG GGAGCGGCGT GCCGT

45

(2) INFORMATION FOR SEQ ID NO:291:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human steroid 5-alpha-reductase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

CTGCCCCCGC GCGCCGCCC TATATGTTGC CCGCCGCGGC CTCTG

45

(2) INFORMATION FOR SEQ ID NO:292:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human substance P receptor gene

358

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

GTGACGTCTC TGCAGGGGGT TATAAAAGCC TCGTGCAGCAG CTAA

44

(2) INFORMATION FOR SEQ ID NO:293:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human synaptobrevin 1 (SYB1) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

CCGGGAGGCG TGCTCAGCAC TAATAAAGGC GGAGGCCGGC GCGGCA

46

(2) INFORMATION FOR SEQ ID NO:294:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human tyrosine aminotransferase (TAT) gene

359

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

CAACGCCCCAT TTGTGGAGAC TATTTCAGGA GTTAGGATTT GCATCTG

47

(2) INFORMATION FOR SEQ ID NO:295:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell receptor V-beta gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

GACAGATGCA TTCTGTGGGG ATAAAATGTC ACAAATTCA TTTCTTT

47

(2) INFORMATION FOR SEQ ID NO:296:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell receptor V-beta gene

360

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

TCACAGAGGG CCTGGTCTAG AATATTCCAC ATCTGCTCTC ACTCT

45

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell receptor V-beta gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

GACAGATGCA TTCTGTGGGG ATAAAATGTC ACAAATTC A TTTCTTT

47

(2) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell receptor V-beta gene

361

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

TCACAGAGGG CCTGGTCTGG AATATTCCAC ATCTGCTCTC ACTCTG

46

(2) INFORMATION FOR SEQ ID NO:299:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell receptor V-beta chain
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

TGTTACTGTA GGAAGTACCG TATAAGGACA GGATGTCCCA CCTCC

45

(2) INFORMATION FOR SEQ ID NO:300:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human transferrin (Tf) gene

362

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

CCGCCCAGGC CGGGAATGGA ATAAAGGGAC GCGGGGCGCC GGAGG

45

(2) INFORMATION FOR SEQ ID NO:301:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interleukin 3 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

GGGCACCTTG

10

(2) INFORMATION FOR SEQ ID NO:302:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human tissue factor gene

363

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

CGGGAGAGCG CGCCGCCGGC CCTTTATAGC GCGCGGGGCA CCGGCTCCCC

50

(2) INFORMATION FOR SEQ ID NO:303:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transforming growth factor-beta (TGF-beta)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

TGCCTTGCCC ATGGGGGCTG TATTTAAGGA CACCGTGCCC CAAGCCC

47

(2) INFORMATION FOR SEQ ID NO:304:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transforming growth factor beta-3 gene

364

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

GAGACGTCAT GGGAGGGAGG TATAAAATT CAGCAGAGAG AAATAGA

47

(2) INFORMATION FOR SEQ ID NO:305:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human transforming growth factor
beta-2 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

CACGTGGTTC AGAGAGAACT TATAAATCTC CCCTCCCCGC GAAGA

45

(2) INFORMATION FOR SEQ ID NO:306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human tyrosine hydroxylase (TH) gene

365

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

GGCTTTGACG TCAGCTCAGC TTATAAGAGG CTGCTGGGCC AGGGCT

46

(2) INFORMATION FOR SEQ ID NO:307:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human metallothionein gene IIA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

TCGTCCCGGC TCTTTCTAGC TATAAACA CTGTTGCCGCG CTGCAC

46

(2) INFORMATION FOR SEQ ID NO:305:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human thrombospondin gene

366

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

CCCAGGAATG CGAGCGCCCC TTAAAAGCG CGCGGCTCCT CCGCCT

46

(2) INFORMATION FOR SEQ ID NO:309:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human thyroxine-binding globulin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

ATAATGTTGC TATAACATCT GAATGACAGT CCATGGCATT ATTTC

45

(2) INFORMATION FOR SEQ ID NO:310:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human thyroglobulin gene

367

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

GAAAGTGCCA ACGGCAGCTC TATAAAGCT CCCTGGCCAG GGGACCT

47

(2) INFORMATION FOR SEQ ID NO:311:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for tumor necrosis factor
(TNF-alpha)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

CTCCTCTCG: CCCAGGGACA TATAAAGGCA GTTGTGGCA CACCCA

46

(2) INFORMATION FOR SEQ ID NO:312:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human lymphotoxin (TNF-beta) gene

368

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

GCTGCCACTG CCGCTTCCTC TATAAAGGGA CCTGAGCGTC CGGGCC

46

(2) INFORMATION FOR SEQ ID NO:313:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human type I DNA topoisomerase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

TGACGTCGCC GACGTGTTGT TTAAAGCGG GCGCGCAGGC GCAGTGAGCC

50

(2) INFORMATION FOR SEQ ID NO:314:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human triosephosphate isomerase (TPI) gene

369

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

AGTTCCACTT CGCGGCGCTC TATATAAGTG GGCAGTGGCC GGACTGC

47

(2) INFORMATION FOR SEQ ID NO:315:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human thyroid peroxidase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

ATCCAAGCGC AGAGTCAGTT TATAAGGTGG GTAACCAAGT CCCT

44

(2) INFORMATION FOR SEQ ID NO:316:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human transferrin receptor gene

370

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

GGCCGGGGGC GGGGCCAGGC TATAAACCGC CGGTTAGGGG CCGCCA

46

(2) INFORMATION FOR SEQ ID NO:317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human tryptase -I gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

CGCCCCCTCC TGATCTGGAA GGATAAATGG GGAGGGGAGA GCCACTGGGT

50

(2) INFORMATION FOR SEQ ID NO:318:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human beta 2 gene for beta-tubulin

371

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:318:

GCGGAGGCGG GCAGGGAGGG TATATAAGCG TTGGCGGACG GTCGGT

46

(2) INFORMATION FOR SEQ ID NO:319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for U 6 RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:319:

GTATTTCGAT TTCTTGGCTT TATATATCTT GTGGAAAGGA CGAAAC

46

(2) INFORMATION FOR SEQ ID NO:320:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human uPA gene for
urokinase-plasminogen activator

372

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:320:

GGCGGCGCCG GGGCGGGCCC TGATATAGAG CAGGCGCCGC GGGTCGC

47

(2) INFORMATION FOR SEQ ID NO:321:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human proto-oncogene vav

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:321:

GCAGGCGTGC GGGCGGGTGG GTGCTGGAGG CTGCCA

36

(2) INFORMATION FOR SEQ ID NO:322:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human vascular cell adhesion
molecule-1 (VCAM1) gene

373

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:322:

GCCTCTGCAA CAAGACCCTT TATAAGCAC AGACTTTCTA TTTCA

45

(2) INFORMATION FOR SEQ ID NO:323:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human vimentin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:323:

ACCCTCTTTC CTAACGGGGT TATAAAAACA GCGCCCTCGG CGGGG

45

(2) INFORMATION FOR SEQ ID NO:324:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human U1 RNA gene

374

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:324:

GTAAAGGGTG AGGTATATGG AGCTGTGACA GGGCAGAAGT GTGTGAAGTC

50

(2) INFORMATION FOR SEQ ID NO:325:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for small nuclear U1 RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:325:

GTAAAGAGTG AGGCGTATGA GGCTGTGTCG GGGCAGAGGC CCAAGATCTC

50

(2) INFORMATION FOR SEQ ID NO:326:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human small nuclear U2 RNA gene

375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:326:

TTGAATGTGG ATGAGAGTGG GACGGTGACG GCGGGCGCGA AGGCGAGCGC

50

(2) INFORMATION FOR SEQ ID NO:327:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human U3 small nuclear RNA gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:327:

AAAAGTTTGC GGCAGATGTA GACCTAGCAG AGGTGTGCGA GGAGGCCGTT

50

(2) INFORMATION FOR SEQ ID NO:328:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human U4C small nuclear RNA gene

376

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:328:

AAATGGTAGT CATCATCCGT GGGGGAGCGG GGCGCGAATA AAGCCTTTCC

50

(2) INFORMATION FOR SEQ ID NO:329:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H3.3 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:329:

GGGCGGGGCG GCGTGTGTTG GGGGATAGCC TCGGTGTCAG CCATCTTTCA

50

(2) INFORMATION FOR SEQ ID NO:330:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human histone H4 gene

377

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:330:

AGTTCGGTCC GCCAACTGTC GTATAAAGGC GCTGCCTCAG GTCAGAGGCC

50

(2) INFORMATION FOR SEQ ID NO:331:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human non-histone chromosomal
protein HMG-14 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:331:

TGGGGGGCGG CCCGGCCGGC GGGGAGGGGG AGCCGCGGCC GGGACGCGGG

50

(2) INFORMATION FOR SEQ ID NO:332:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human ribosomal protein S14 gene

378

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:332:

AAGTAATAAA CCGTCTTTCC TTATGACGAG TCTTAACTC TTTGGGAGGA

50

(2) INFORMATION FOR SEQ ID NO:333:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for alpha tubulin (b
alpha 1)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:333:

CGCGACCGAG GGTCTGGGCG TCCCGGCTGG GCCCCGTGTC TGTGCGCAG

50

(2) INFORMATION FOR SEQ ID NO:334:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human skeletal alpha-actin gene

379

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:334:

AGGGAATCGC CCGCGGGCTA TATAAACCT GAGCAGAGGG ACAAGCGGCC

50

(2) INFORMATION FOR SEQ ID NO:335:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human epidermal 67-kDa Keratin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

GGAAGATCTT GTGTGATAAA ACAATTACCA CATGAACCA TCTTGCATGC

50

(2) INFORMATION FOR SEQ ID NO:336:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human 50 KDa type I epidermal keratin gene

380

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:336:

GACCCGCCCC CTACCCATGA GTATAAGCA CTCGCATCCC TTTGCAATTT

50

(2) INFORMATION FOR SEQ ID NO:337:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-1 collagen type I gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

CTGCTCTCCA TCAGGACAGT ATAAAAGGGG CCCGGGCCAG TCGTCGGAGC

50

(2) INFORMATION FOR SEQ ID NO:338:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human collagen type-III gene

381

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:338:

GTGAGGGAAG CCAAACCTTTT TCCTATTTAA GGCCAAAGCA AAGGAATCTC

50

(2) INFORMATION FOR SEQ ID NO:339:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pro-alpha-2 (I) mRNA for
collagen N-prepropeptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:339:

CAGGGAAACT TTTGCCGTAT AAATAGGGCA CATCCGGGAT TTGTTATTTT

50

(2) INFORMATION FOR SEQ ID NO:340:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human fibronectin (FN) gene

382

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:340:

TCCAGAGGGG CGGGAGGGCC GTCCCATATA AGCCCGGCTC CCGCGCTCCG

50

(2) INFORMATION FOR SEQ ID NO:341:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human von Willebrand factor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:341:

TGTTTCCTTT TGGTAATTAA AAGGAGGCCA ATCCCCTGTT GTGGCAGCTC

50

(2) INFORMATION FOR SEQ ID NO:342:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma chain

383

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:342:

TGGCTATCCC AGGAGCTTAC ATAAAGGGAC AATTGGAGCC TGAGA

45

(2) INFORMATION FOR SEQ ID NO:343:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma chain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

CAGTCCTGGC TATCCCAGGA GCTTACATAA AGGGACAATT GGAGCCTGAG

50

(2) INFORMATION FOR SEQ ID NO:344:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human involucrin mRNA

384

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

AGGCCAGGCT GCAGAATGAT ATAAAGAGTG CCCTGACTCC TGCTCAGCTC

50

(2) INFORMATION FOR SEQ ID NO:345:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human apolipoprotein A-I and C-III genes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

CCAGACCCTG GCTGCAGACA TAAATAGGCC CTGCAAGAGC TGGCTGCTTA

50

(2) INFORMATION FOR SEQ ID NO:346:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human apolipoprotein B-100 (apoB) gene

385

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

GCTCTTGCAG CCTGGGCTTC CTATAAATGG GGTGCGGGCG CCGGCCGCGC

50

(2) INFORMATION FOR SEQ ID NO:347:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human apolipoprotein A-I and C-III genes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:347:

TCTAGGGATG AACTGAGCAG

20

(2) INFORMATION FOR SEQ ID NO:348:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Humanapolipoprotein A-I and C-III genes

386

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:348:

ACAGGCAGGA GGTTCTGAC CTGTTTATA TCATCTCCAG GGCAGCAGGC A

51

(2) INFORMATION FOR SEQ ID NO:349:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human albumin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:349:

TACAATTATT GGTAAAGAA GTATATTAGT GCTAATTTC CTCCGTTTGT

50

(2) INFORMATION FOR SEQ ID NO:350:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human albumin gene

387

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:350:

TACAATTATT GGTAAAGAA GTATATTAGT GCTAATTTCC CTCCGTTTGT C

51

(2) INFORMATION FOR SEQ ID NO:351:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human serum prealbumin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:351:

CCTAGCTCAG GAGAAGTGAG TATAAAAGCC CCAGGCTGGG AGCAGCCATC

50

(2) INFORMATION FOR SEQ ID NO:352:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-fetoprotein (AFP) gene

388

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:352:

TAACAGGCAT TGCCTGAAAA GAGTATAAAA GAATTCAGC ATGATTTTCC

50

(2) INFORMATION FOR SEQ ID NO:353:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human C-reactive protein gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:353:

AGGCAGGAGG AGGTAGCTCT AAGGCAAGAG ATCTGGGACT T

41

(2) INFORMATION FOR SEQ ID NO:354:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene A for alpha 1-acid
glycoprotein

389

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:354:

AAGTGACCGC CCATAGTTTA TTATAAAGGT GACTGCACCC TGCAGCCACC

50

(2) INFORMATION FOR SEQ ID NO:355:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene A for alpha 1-acid glycoprotein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:355:

AAGTGACCGC CCATAGTTTA TTATAAAGGT GACTGCACCC TGCAGCCACC A

51

(2) INFORMATION FOR SEQ ID NO:356:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for L apoferritin

390

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:356:

CGGCGCACCA TAAAGAAGC CGCCCTAGCC ACGTCCCCTC

40

(2) INFORMATION FOR SEQ ID NO:357:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for L apoferritin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:357:

CGGCGCACCA TAAAGAAGC CGCCCTAGCC ACGTCCCCTC G

41

(2) INFORMATION FOR SEQ ID NO:358:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Olive baboon alpha-1 globin gene

391

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:358:

GGCGTGCCCC CGCGCCCGGA GCATAAACCC TGGCGCGCTC GCGGCCCGGC

50

(2) INFORMATION FOR SEQ ID NO:359:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Olive baboon alpha-1 globin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:359:

GGCGTGCCCC CGCGCCCGGA GCATAAACCC TGGCGCGCTC GCGGCCCGGC A

51

(2) INFORMATION FOR SEQ ID NO:360:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-globin germ line gene

392

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:360:

GTGCCAACAA TGGAGGTGTT TACCTGTCTC AGACCAAGGA CCTCTCTGCA

50

(2) INFORMATION FOR SEQ ID NO:361:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Chimpanzee gene for alpha-like
zeta-1-globin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:361:

CTGGCTGGG CCCAGCTCCC TGTATATAAG GGGACCCTGG GGGCTGAGCA

50

(2) INFORMATION FOR SEQ ID NO:362:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Chimpanzee gene for alpha-like
zeta-1-globin

393

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:362:

CCTGGCTGGG CCCAGCTCCC TGTATATAAG GGGACCCTGG GGGCTGAGCA C

51

(2) INFORMATION FOR SEQ ID NO:363:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human alpha globin gene cluster on chromosome 16: zeta gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:363:

CTGGCTGGGC CCAGCTCCCT GTATATAAGG GGACCCTGGG GGCTGAGCAC

50

(2) INFORMATION FOR SEQ ID NO:364:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human theta 1-globin gene

394

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:364:

CCGCGGGACC CCTGGCCGGT CCGCGCAGGC GCAGCGGGGT CGCAGGGCGC

50

(2) INFORMATION FOR SEQ ID NO:365:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Macaque cynomolgus beta-globin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:365:

GCAGGAGCCA GGGCTGGGCA TAAAAGTCAG GGCAGAGCCA TCTATTGCTT

50

(2) INFORMATION FOR SEQ ID NO:366:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Chimpanzee beta-globin gene

395

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:366:

GCAGAAGCCA GGGCTGGGCA TAAAGTCAG GGCAGAGCCA TCTATTGCTT

50

(2) INFORMATION FOR SEQ ID NO:367:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human germ line gene for beta-globin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:367:

GCAGGAGCCA GGGCTGGGCA TAAAGTCAG GGCAGAGCCA TCTATTGCTT

50

(2) INFORMATION FOR SEQ ID NO:368:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Spider monkey (A.geoffroyi)
delta-globin gene

396

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:368:

CAGGGAGAAC AGGACCAGCA TAAAGGCAG GGCAGGGCTA ACTGTTGCTT

50

(2) INFORMATION FOR SEQ ID NO:369:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human transferrin receptor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:369:

GGGCGGGGCC AGGCTATAAA CCGCCGGTTA GGGGCCGCCA TCCCTCAGA

50

(2) INFORMATION FOR SEQ ID NO:370:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human beta-2-adrenergic receptor gene

397

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:370:

AGTTCCCCTA AAGTCCTGTG CACATAACGG GCAGAACGCA CTGCGAAGCG

50

(2) INFORMATION FOR SEQ ID NO:371:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human IgE receptor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:371:

GGTGGCAAGC CCATATTTAG GTCTATGAAA ATAGAAGCTG TCAGTGGCTC

50

(2) INFORMATION FOR SEQ ID NO:372:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human oncogene c-fos

398

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:372:

TTCATAAAAC GCTTGTTATA AAAGCAGTGG CTGCGGCGCC TCGTACTCCA

50

(2) INFORMATION FOR SEQ ID NO:373:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human c-myc oncogene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:373:

AATCTCCGCC CACCGGCCCT TTATAATGCG AGGGTCTGGA CGGCTGAGGA

50

(2) INFORMATION FOR SEQ ID NO:374:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human B-cell leukemia/lymphoma 2
(bcl-2) proto-oncogene

399

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:374:

CCGCCCCCTCC GCGCCGCCTG CCCGCCCGCC CGCCGCGCTC CCGCCCCGCCG

50

(2) INFORMATION FOR SEQ ID NO:375:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human p53 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:375:

ACTCCATTTC CTTTGCTTCC TCCGGCAGGC GGATTACTTG CCCTTACTTG

50

(2) INFORMATION FOR SEQ ID NO:376:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene homologous to bladder carcinoma oncogene T24

400

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:376:

CGCGGGCCCTA CTGGCTCCGC CTCCCGCGTT GCTCCCGGAA GCCCCGCCCCG

50

(2) INFORMATION FOR SEQ ID NO:377:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human c-abl gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:377:

GGGGCGGGCC TGGCGGGCGC CCTCTCCGGG CCCTTTGTTA ACAGGCGCGT

50

(2) INFORMATION FOR SEQ ID NO:378:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human metallothionein-i-a gene

401

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:378:

CGGCCCTCTT TCCCCTGACC ATAAAAGCAG CCGCTGGCTG CTGGGCCCTA

50

(2) INFORMATION FOR SEQ ID NO:379:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human metallothionein I-B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:379:

ACCCACACCAC CTCCCCGAC TATAAAGGAG CAGCCAGCTC CTGGGCTCCA

50

(2) INFORMATION FOR SEQ ID NO:380:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human metallothionein-If (MT-IF)
gene

402

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:380:

CCCGGCCCCC TCCCCTGACT ATCAAAGCAG CGGCCGGCTG TTTGGGTCCA

50

(2) INFORMATION FOR SEQ ID NO:381:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for 27 Kda heat shock protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:381:

CCCTCAAACG GGTCAATGCC ATTAATAGAG ACCTCAAACA CCGCCTGCTA

50

(2) INFORMATION FOR SEQ ID NO:382:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human 70 kDa heat shock protein gene

403

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:382:

GCGGGTCTCC GTGACGACTT ATAAACCCC AGGGGCAAGC GGTCCGGATA

50

(2) INFORMATION FOR SEQ ID NO:383:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human macrophage alaph-antitrypsin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:383:

TGCCTCCACC CGAAGTCTAC TTCCTGGGTG GGCAGGAAGT GGGCACTGTG

50

(2) INFORMATION FOR SEQ ID NO:384:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human alaph-antitrypsin (S variant) gene

404

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:384:

CGTTGCCCCCT CTGGATCCAC TGCTTAAATA CGGACGAGGA CAGGGCCCTG

50

(2) INFORMATION FOR SEQ ID NO:385:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human S variable segment 5' of
antithrombin III gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:385:

TCTGCCCCAC CCTGTCCTCT GGAACCTCTG CGAGATTTAG AGGAAAGAAC

50

(2) INFORMATION FOR SEQ ID NO:386:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human pulmonary surfactant protein
(SP5) gene

405

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:386:

CCCCTCTCCC TACGGACACA TATAAGACCC TGGTCACACC TGGGAGAGGA

50

(2) INFORMATION FOR SEQ ID NO:387:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human Immunoglobulin kappa L-chain V region gene (HK122)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:387:

CCCCCTGCCC TGAAGACTTT TTATAGGCTG GTCACACCCG GAGCAGGAGT

50

(2) INFORMATION FOR SEQ ID NO:388:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human T cell receptor V-alpha/J-alpha chain (rearranged)

406

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:388:

TTAAGGTTTG AATCCTCAGT GAACCAGGGC AGAAAAGAAT GATGAAATCC

50

(2) INFORMATION FOR SEQ ID NO:389:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for HLA-DR alpha heavy chain (class II antigen)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:389:

TGCATTTTAA TGGTCAGACT CTATTACACC CCACATTCTC TTTTCTTTTA

50

(2) INFORMATION FOR SEQ ID NO:390:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human MHC class IIHLA-DC-3-beta gene

407

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:390:

CTACCACGCA TGGAAACATC CACAGATTTT TATTCTTTCT GCCAGGTACA

50

(2) INFORMATION FOR SEQ ID NO:391:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell receptor CD3-gamma gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:391:

GCCTTCTCTC AAAGGCCCCA GCCCCAACAG TGATGGGTGG AGCCAGTCTA

50

(2) INFORMATION FOR SEQ ID NO:392:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pregnancy-specific beta-1
glycoprotein mRNA

408

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:392:

CTGCCCTGGG AAGAGGCTCA GCACAGAAAG AGGAAGGACA GCACAGCTGA

50

(2) INFORMATION FOR SEQ ID NO:393:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human pregnancy-specific
beta-1-glycoprotein 5 (PSG5)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:393:

AGAGAGGAGG GGACAGAGAG GTGTCCTGGG CCTGACCCCA CCCATGAGCC

50

(2) INFORMATION FOR SEQ ID NO:394:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human factor VIII gene

409

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:394:

CCTGTGGCTG CTTCCCACTG ATAAAAAGGA AGCAATCCTA TCGGTTACTG

50

(2) INFORMATION FOR SEQ ID NO:395:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ubiquitin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:395:

TGACGCAACA CTCGTTGCAT AAATTTGCCT CCGCCAGCCC GGAGCATTTA

50

(2) INFORMATION FOR SEQ ID NO:396:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human proliferating cell nucleolar
protein P120 gene

410

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:396:

ACTATAATAC GCCAAGCGTG CGTTCTGCCG TTCCCTCCGA CACGCGCGAC

50

(2) INFORMATION FOR SEQ ID NO:397:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for delta-globin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:397:

CAGGGAGGAC AGGACCAGCA TAAAAGGCAG GGCAGAGTCG ACTGTTGCTT

50

(2) INFORMATION FOR SEQ ID NO:398:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Gorilla fetal A-gamma-globin gene

411

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:398:

CGGCTGGCTA GGGATGAAGA ATAAAAGGAA GCACCCTCCA GCAGTTCAC

50

(2) INFORMATION FOR SEQ ID NO:399:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for fetal A-gamma and G-gamma hemoglobin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:399:

CGGCTGGCTA GGGATGAAGA ATAAAAGGAA GCACCCTTCA GCAGTTCAC

50

(2) INFORMATION FOR SEQ ID NO:400:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Orangutan epsilon-globin gene with flanking Alu repeats

412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:400:

CAGAACTTCG GCAGTGAAGA ATAAAAGGCC ACACAGAGAG GCAGCAGCAC

50

(2) INFORMATION FOR SEQ ID NO:401:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human haptoglobin (Hpl)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:401:

TAAAAGACC AGCAGATGCC CCACAGCACT GCTCTTCCAG AGGCAAGACC

50

(2) INFORMATION FOR SEQ ID NO:402:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human low molecular weight
oligoadenylate synthetase gene

413

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:402:

AAGACAGCTC CTCCCTTCTG AGGAAACGAA ACCAACAGCA GTCCAAGCTC AGTCAGCAGA 60
AGAGATAAAA G 71

(2) INFORMATION FOR SEQ ID NO:403:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene fragment for dihydrofolate reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:403:

GGGGGGCGGG GCCTCGCCTG CACAAATAGG GACGAGGGGG CGGGGCGGCC 50

(2) INFORMATION FOR SEQ ID NO:404:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

414

(C) INDIVIDUAL ISOLATE: Human thymidine kinase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:404:

GGCTCGTGAT TGGCCAGCAC GCCGTGGTTT AAAGCGGTCG GCGCGGGACC

50

(2) INFORMATION FOR SEQ ID NO:405:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human adenosine deaminase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:405:

GCGGGAGGCG GGGCCCGGCC CGTTAAGAAG AGCGTGGCCG GCCGCGGCC

49

(2) INFORMATION FOR SEQ ID NO:406:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human argininosuccinate synthase

415

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:406:

TGCCCCCGGG CCCTGTGCTT ATAACCTGGG ATGGGCACCC CTGCCAGTCC

50

(2) INFORMATION FOR SEQ ID NO:407:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human ornithine aminotransferase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:407:

GGGGGCGGGG CAGAATCAGC CTTTAAGTTG CAGTGACGCT CCGGCGTCAC

50

(2) INFORMATION FOR SEQ ID NO:408:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human tyrosine hydroxylase gene

416

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:408:

TGACGTCAGC TCAGCTTATA AGAGGCTGCT GGGCCAGGGC TGTGG

45

(2) INFORMATION FOR SEQ ID NO:409:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human HMG CoA reductase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:409:

CAGCTCCGAG CGTGCCTAAG GTGAGGGCTC CTTCCGCTCC GCGACTGCGT

50

(2) INFORMATION FOR SEQ ID NO:410:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for lecithin-cholesterol
acyltransferase

417

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:410:

CCTAGGGCCC CTCCCACTCC CACACCAGAT AAGGACAGCC CAGTGCCCGCT

50

(2) INFORMATION FOR SEQ ID NO:411:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human porphobilinogen deaminase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:411:

CGCCCAGAGG GAGGGACCTC CCCTTCGAGG GAGGGCGCCG GAAGTGACGC

50

(2) INFORMATION FOR SEQ ID NO:412:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human porphobilinogen deaminase gene

418

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:412:

GCACAGCACT CCCACTGACA ACTGCCTTGG TCAAGGTGGG CTCAGGGCT

50

(2) INFORMATION FOR SEQ ID NO:413:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human URO-D gene for
uroporphyrinogen decarboxylase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:413:

GGGGGGCAGG CTCAGATTCA GGTTAAATTG TGGATTGAGC TCGCAGTTAC

50

(2) INFORMATION FOR SEQ ID NO:414:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human URO-D gene for
uroporphyrinogen decarboxylase

419

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:414:

GGGGGGCAGG CTCAGATTC A GGTAAATTG TGGATTGAGC TCGCAGTTAC A

51

(2) INFORMATION FOR SEQ ID NO:415:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase B (ALDOB) gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:415:

AAAAAAAAA CATGATGAGA AGTCTATAAA AATTGTGTGC TACCAAAGAT

50

(2) INFORMATION FOR SEQ ID NO:416:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

420

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:416:

GGTGGCGCTG CTCACCACAC ACAAGTGTTA TAGGAGGAGT CTGGCCCTTG

50

(2) INFORMATION FOR SEQ ID NO:417:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:417:

GGTGGCGCTG CTCACCACAC ACAAGTGTTA TAGGAGGAGT CTGGCCCTTG A

51

(2) INFORMATION FOR SEQ ID NO:418:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

421

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:418:

TGTGGGGCGG GCAGGAGCTG CCTTATAACC AGCCCGGGAA CCCCTAGCTC

50

(2) INFORMATION FOR SEQ ID NO:419:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:419:

TGTGGGGCGG GCAGGAGCTG CCTTATAACC AGCCCGGGAA CCCCTAGCTC A

51

(2) INFORMATION FOR SEQ ID NO:420:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

422

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:420:

GCTCGGCGGA GGGCGGAGTG GTGCCTTTAA AAGGCCGGGC GCCGCCTTCC

50

(2) INFORMATION FOR SEQ ID NO:421:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:421:

GCTCGGCGGA GGGCGGAGTG GTGCCTTTAA AAGGCCGGGC GCCGCCTTCC G

51

(2) INFORMATION FOR SEQ ID NO:422:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

423

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:422:

GCTAAATCGG CTGCGTTCCT CTCGGAACGC GCCGCAGAAG GGGTCCTGGT

50

(2) INFORMATION FOR SEQ ID NO:423:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human aldolase A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:423:

GCTAAATCGG CTGCGTTCCT CTCGGAACGC GCCGCAGAAG GGGTCCTGGT G

51

(2) INFORMATION FOR SEQ ID NO:424:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human phosphoglycerate kinase gene

424

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:424:

GAGGCGGGGT GTGGGGCGGT AGTGTGGGCC CTGTTCTGTC CCGCGCGGTG

50

(2) INFORMATION FOR SEQ ID NO:425:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for glucose 6-phosphate
dehydrogenase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:425:

CAGGCGCCCG CCCCCGCCCC CGCCGATTAA ATGGGCCGGC GGGGCTCAGC

50

(2) INFORMATION FOR SEQ ID NO:426:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hepatic lipase gene

425

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:426:

GCAGTCTTCC CTAACAAAGT ATCTAATAGG CATTGTGGTC TCTTTGGCTT

50

(2) INFORMATION FOR SEQ ID NO:427:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human hepatic lipase mRNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:427:

GCAGTCTTCC CTAACAAAGT ATCTAATAGG CATTGTGGTC TCTTTGGCTT C

51

(2) INFORMATION FOR SEQ ID NO:428:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human protein C gene

426

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:428:

AGTGCTGAGG GCCAAGCAAA TATTGTGGT TATGGATTAA CTCGAACTCC

50 -

(2) INFORMATION FOR SEQ ID NO:429:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human factor IX gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:429:

CCAGAAGTAA ATACAGCTCA GCTTGACTT TGGTACAACT AATCGACCTT

50

(2) INFORMATION FOR SEQ ID NO:430:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human MHC III HLA factor B gene

427

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:430:

GCAGGTGCCA GAACACAGAT TGTATAAAAG GCTGGGGGCT GGTGGGGAGC

50

(2) INFORMATION FOR SEQ ID NO:431:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pepsinogen gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:431:

CGATAAGGCG GGACCCAACT TGTATATAAG GGCAGCTCAT GCTGCTGCTC

50

(2) INFORMATION FOR SEQ ID NO:432:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pepsinogen C gene

428

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:432:

CGATTAGACT AATCTTGGGC GTATAAAGA GGAAAGAGTG CCCAGGTCTT

50

(2) INFORMATION FOR SEQ ID NO:433:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human collagenase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:433:

CTGGAAGGGC AAGGACTCTA TATATACAGA GGGAGCTTCC TAGCTGGGAT

50

(2) INFORMATION FOR SEQ ID NO:434:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human stromelysin gene

429

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:434:

CCAAACAAAC ACTGTCACTC TTAAAAAGCT GCGCTCCCGA GGTGGACCT

50

(2) INFORMATION FOR SEQ ID NO:435:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human alpha-amylase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:435:

TCTGATCCGT GCAGGGTATT AATGTGTCAG GGCTGAGTGT TCTGAGATT

50

(2) INFORMATION FOR SEQ ID NO:436:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pancreatic alpha-amylase gene

430

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:436:

TGTAAATGT GCTTCTTACA GGAATATAAA TAGTTTCTGG AAAGGACACT

50

(2) INFORMATION FOR SEQ ID NO:437:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human pancreatic amylase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:437:

TGTAAATGT GCTTCTTACA GGAATATAAA TAGTTTCTGG AAAGGACACT

50

(2) INFORMATION FOR SEQ ID NO:438:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: human cytochrome P450c gene

431

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:438:

GCCACACGTA CAAGCCCGCC TATAAAGGTG GCAGTGCCTT CACCCTCACC

50

(2) INFORMATION FOR SEQ ID NO:439:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytochrome P-450c gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:439:

GCCACACGTA CAAGCCCGCC TATAAAGGTG GCAGTGCCTT CACCCTCACC C

51

(2) INFORMATION FOR SEQ ID NO:440:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for cytochrome P(1)-450

432

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:440:

CACGTACAAG CCCGCCTATA AAGGTGGCAG TGCCTTCACC

40

(2) INFORMATION FOR SEQ ID NO:441:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human steroid 21-hydroxylase [P450
(C21)] B gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:441:

GGATGGCTGG GGCTCTTGAG CTATAAGTGG CACCTCAGGG CCCTGACGGG

50

(2) INFORMATION FOR SEQ ID NO:442:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human mitochondrial aldehyde
dehydrogenase 2 gene

433

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:442:

TTCCTGACCA TGGTACTTAT AAAAGCAGTG CCGTCTGCCC CATCCATGTC

50

(2) INFORMATION FOR SEQ ID NO:443:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human carbonic anhydrase III gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:443:

AAGGCCATGC AAGTGTGCGG GGGAGCTACA TAAAGCCGG GGCTCGCGCG

50

(2) INFORMATION FOR SEQ ID NO:444:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human creatine kinase B isozyme gene

434

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:444:

TGGGCGGCCC GCGTTGTGCC CCTTAAGAGC CGCGGGAGCG CGGAGCGGCC

50

(2) INFORMATION FOR SEQ ID NO:445:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human preproenkephalin A gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:445:

CTTCGGTTTG GGGCTAATTA TAAAGTGGCT CCAGCAGCCG TTAAGCCCCG

50

(2) INFORMATION FOR SEQ ID NO:446:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human preprokephalin A gene

435

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:446:

CTTCGGTTTG GGGCTAATTA TAAAGTGGCT CCAGCAGCCG TTAAGCCCCG G

51

(2) INFORMATION FOR SEQ ID NO:447:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human prepro form of corticotropin releasing factor gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:447:

TTTTTGAAGA GGGTCGACAC TATAAAATCC CACTCCAGGC TCTGGAGTGG

50

(2) INFORMATION FOR SEQ ID NO:448:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human preprothyrotropin-releasing hormone gene

436

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:448:

GACCTCACTC GAGCCGCCGC CTGGCGCAGA TATAAGCGGC GGCCCATCTG

50

(2) INFORMATION FOR SEQ ID NO:449:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for fetal A-gamma and G-gamma hemoglobin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:449:

CGGCTGGCTA GGGATGAAGA ATAAAAGGAA GCACCCTTCA GCAGTTCCAC

50

(2) INFORMATION FOR SEQ ID NO:450:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene coding for ACTH and beta-LPH precursors

437

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:450:

CCCACCAGGA GAGCTCGGCA AGTATATAAG GACAGAGGAG CGCGGGACCA

50

(2) INFORMATION FOR SEQ ID NO:451:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human somatostatin I gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:451:

TAGCCTGACG TCAGAGAGAG AGTTTAAAC ACAGGGAGAC GGTGAGAGC

50

(2) INFORMATION FOR SEQ ID NO:452:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human glucagon gene

438

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:452:

GTGAGGCTAA ACAGAGCTGC AGAGTATATA AAAGCAGTGC GCCTTGGTGC

50

(2) INFORMATION FOR SEQ ID NO:453:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human glucagon gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:453:

GTGAGGCTAA ACAGAGCTGG AGAGTATATA AAAGCAGTGC GCCTTGGTGC A

51

(2) INFORMATION FOR SEQ ID NO:454:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human chorionic gonadotropin gene

439

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:454:

AGGTGGAAAC ACTCTGCTGG TATAAAAGCA GGTGAGGACT TCATTAAGTG

50

(2) INFORMATION FOR SEQ ID NO:455:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human chorionic gonadotropin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:455:

TTGAACTGTG GTGCAGGAAA GCCTCAAGTA GAGGAGGGTT GAGGCTTCAA

50

(2) INFORMATION FOR SEQ ID NO:456:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human beta-LH gene

440

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:456:

GCCGCCCCCA CAACCCCGAG GTATAAGCC AGATACACGA GGCAGGGGAT

50

(2) INFORMATION FOR SEQ ID NO:457:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human follicle-stimulating hormone
beta-subunit gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:457:

TAGTTGCACA TGATTTTGTA TAAAAGGTGA ACTGAGATT CATTCACTCT

50

(2) INFORMATION FOR SEQ ID NO:458:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human prolactin gene

441

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:458:

TATTCATGAA GATATCAAAG GTTATAAAG CCAATATCTG GGAAAGAGAA

50

(2) INFORMATION FOR SEQ ID NO:459:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human parathyroid gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:459:

GACATCATCT GTAACAATAA AAGAGCCTCT CTTGGAAGC AGAAGACCTA

50

(2) INFORMATION FOR SEQ ID NO:460:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Owl monkey insulin gene

442

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:460:

GGGGAGATGG GCTCTGGGCC TATAAGCCA GCAGGGACCC AGCAGCCCTC

50

(2) INFORMATION FOR SEQ ID NO:461:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: human insulin/IGF II gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:461:

CCCCGCCTCC AGAGTGGGGG CCAAGGCTGG GCAGGCGGGT GGACGGCCGG

50

(2) INFORMATION FOR SEQ ID NO:462:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human insulin like growth factor
IGFII gene

443

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:462:

AAAGAACTCT GCCTTGC GTT CCCCAAATT TGGGCATTGT TCCGGCTCGC

50

(2) INFORMATION FOR SEQ ID NO:463:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human insulin-like growth factor II gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:463:

CCCTGGGCCG CGGCTGGCGC GACTATAAGA GCCGGGCGTG GCGCCCCGCA

50

(2) INFORMATION FOR SEQ ID NO:464:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gastrin gene

444

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:464:

AGTTGGGAGG GACCTTGAGG GCTTTATAAG GCAGGCCTGG AGCATCAAGC

50

(2) INFORMATION FOR SEQ ID NO:465:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
INF-alpha 13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:465:

GGAAATCAGT ATGTTCCCTA TTTAAGGCAT CTGCAGGAAG CAAAGCCTTC

50

(2) INFORMATION FOR SEQ ID NO:466:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for leukocyte interferon

445

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:466:

GGAAGCTAGT ATGTTCTTA TTTAAGACCT ATGCACAGAG CAAGGTCTTC

50

(2) INFORMATION FOR SEQ ID NO:467:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon alpha gene
INF-alpha 4b

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:467:

GGAAATTAGT ATGTTCACTA TTTAAGACCT ATGCACAGAG CAAAGTCTTC

50

(2) INFORMATION FOR SEQ ID NO:468:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human gene for leukocyte (alpha)
interferon

446

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:468:

GGAAATTAGT ATGTTCACTA TTTAAGGCCT ATGCACAGAG CAAAGTCTTC

50

(2) INFORMATION FOR SEQ ID NO:469:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human interferon genes LeIF-L and
LeIF-J

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:469:

GGAAATTAGT ATGTTCACTA TTTAAGACCT ATGCACAGAG CAAAGTCTTC

50

(2) INFORMATION FOR SEQ ID NO:470:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human gene for fibroblast (beta-1)
interferon

447

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:470:

ATAGAGAGAG GACCATCTCA TATAAATAGG CCATACCCAC GGAGAAAGGA

50

(2) INFORMATION FOR SEQ ID NO:471:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human c-sis gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:471:

CTCTCGCACT CTCCCTTCTC CTTTATAAAG GCCGGAACAG CTGAAAGGGT

50

(2) INFORMATION FOR SEQ ID NO:472:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human migratory inhibitory
factor-related protein 8

448

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:472:

CAGCTGGCCA AGCCTAACCG CTATAAAAAG GAGCTGCCTC TCAGCCCTGC

50

(2) INFORMATION FOR SEQ ID NO:473:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human migratory inhibitory
factor-related protein 14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:473:

GTGCCCCAGT CAGGAGCTGC CTATAAATGC CGAGCCTGCA CAGCTCTGGC

50

(2) INFORMATION FOR SEQ ID NO:474:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human epidermal growth factor
related gene

449

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:474:

GGTCCCTCCT CCTCCGCCC TGCCTCCCGC GCCTCGGCCC GCGCGAGCTA

50

(2) INFORMATION FOR SEQ ID NO:475:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human opsin gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:475:

GCTTAGGAGG GGGAGGTCAC TTTATAAGGG TCTGGGGGGG TCAGAACCCA

50

(2) INFORMATION FOR SEQ ID NO:476:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human blue cone photoreceptor
pigment gene

450

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:476:

TTTTGTGGGG TGGGAGGATC ACCTATAAGA GGA CTCA GAG GAGGGTGTGG

50

(2) INFORMATION FOR SEQ ID NO:477:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human red cone photoreceptor pigment gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:477:

CGGGCTGATC CCACAGGCCA GTATAAAGCG CCGTGACCCT CAGGTGATGC

50

(2) INFORMATION FOR SEQ ID NO:478:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human green cone photoreceptor pigment gene

451

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:478:

CGGGCTGATC CCACTGGCCG GTATAAGCG CCGTGACCCT CAGGTGACGC

50

(2) INFORMATION FOR SEQ ID NO:479:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon-inducible gene
IFI-56K

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:479:

TTGGCTGCTG TTTAGCTCCC TTATATAACA CTGTCTTGGG GTTTAAACGT

50

(2) INFORMATION FOR SEQ ID NO:480:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human interferon-induced 15-Kd
protein (ISG) gene

452

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:480:

GACGTGTGTG CCTCAGGCTT AATAATAGGG CCGGTGCTGC TCGGAAGCC

50

(2) INFORMATION FOR SEQ ID NO:481:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human ubiquitin-like protein (GdX)
gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:481:

TCCAGCGCGC GCGCCCGGGG CGGCGGCGCG CGGCGGGGGG TGTTGGGGT

50

(2) INFORMATION FOR SEQ ID NO:482:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human exogenous retrovirus erv3 5"
long terminal repeat

453

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:482:

CCGCCCCCTGT TGGTTGCATG TATAAAGTC AAGCCCTGTC ATTGTTTCAGG

50

(2) INFORMATION FOR SEQ ID NO:483:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Bovine leukemia virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:483:

ACCTCACCTG CTGATAAATT AATAAAATGC CGGCCCTGTC GAGTTAGCGG

50

(2) INFORMATION FOR SEQ ID NO:484:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human T-cell lymphotropic virus type

I

454

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:484:

TCAATAAACT AGCAGGAGTC TATAAAAGCG TGGAGACAGT TCAGGAGGGG

50

(2) INFORMATION FOR SEQ ID NO:485:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human T-cell leukemia virus II
proviral LTR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:485:

TCAAAATAAA AGATGCCGAG TCTATAAAAG CGCAAGGACA GTTCAGGAGG

50

(2) INFORMATION FOR SEQ ID NO:486:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human T-cell Lymphotropic virus type
III (HIV-1)

455

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:486:

GGCGAGCCCT CAGATCCTGC ATATAAGCAG CTGCTTTTGG CCTGTACTGG

50

(2) INFORMATION FOR SEQ ID NO:487:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Aids-associated retrovirus
(arv-2;proviral)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:487:

TGGCGTCCCT CAGATGCTGC ATATAAGCAG CTGCTTTTGG CCTGTACTGG

50

(2) INFORMATION FOR SEQ ID NO:488:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human immunodeficiency virus type 2
(HIV-2)

456

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:488:

GCCCTCATAT TCTCTGTATA AATATACCCG CTAGCTTGCA TTGTACTTCG

50

(2) INFORMATION FOR SEQ ID NO:489:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Visna lentivirus, Icelandic strains
LV1-1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:489:

CATAACCGCA GATGTAAACA AGTTGCCTAT ATAAGCCGCT TGCTAGCTGG

50

(2) INFORMATION FOR SEQ ID NO:490:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus strain AD169
gene I

457

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:490:

GGCGTGACG GTGGGAGGTC TATATAAGCA GAGCTCGTTT AGTGAACCGT

50

(2) INFORMATION FOR SEQ ID NO:491:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Murine cytomegalovirus
intermediate-early gene I

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:491:

GCTGAGCTGC GTTCACGTGG GTATAAGAGG CGCGACCAGC GTCGGTACCG

50

(2) INFORMATION FOR SEQ ID NO:492:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus
intermediate-early glycoprotein UL37

458

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:492:

CGTCATGTCC GGCATCTTCA TGTATATAAG ACGGTGTTTC AAGACGACGT

50

(2) INFORMATION FOR SEQ ID NO:493:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytomegalovirus I-E
glycoprotein US3 gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:493:

ACAACGTCAC CAAGAAACGC TATATATTCA AAAACACCGT TCAGTCCACA

50

(2) INFORMATION FOR SEQ ID NO:494:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Herpes Simplex Virus type 1 gene I

459

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:494:

TTTGSGGAGG GGAAAGGCGT GGGGTATAAG TTAGCCCTGG CCCGACAGTC

50

(2) INFORMATION FOR SEQ ID NO:495:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Herpes Simplex Virus type 1 gene II

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:495:

AGCCGGCCCC GGCACCACGG GTATAAGGAC ATCCACCACC CGGCCGGTGG

50

(2) INFORMATION FOR SEQ ID NO:496:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Herpes simplex virus type II I-E
gene II

460

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:496:

AGCCGGCCCC GGTCGTGCGG GTATAAGGGC AGCCACCGGC CCACTGGGCG

50

(2) INFORMATION FOR SEQ ID NO:497:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus type I I-E gene
III

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:497:

TTCCCGCCGG CCCCTGGGAC TATATGAGCC CGAGGACGCC CCGATCGTCC

50

(2) INFORMATION FOR SEQ ID NO:498:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus type II I-E
gene III

461

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:498:

CCCCGCGCGC CCCGAGCGAC TATATCAGCC AGGCGACGGG GCGATCGTCC

50

(2) INFORMATION FOR SEQ ID NO:499:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus type 1 I-E
genes IV and V

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:499:

GGGGGCGGGT CTCTCCGGCG CACATAAAGG CCCGGCGCGA CCGACGCCCCG

50

(2) INFORMATION FOR SEQ ID NO:500:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus type 2 I-E
genes IV and V

462

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:500:

ACGGGGGGCG GGCCGTTTCCT CGCGCACATA AAGGGCCGGC GTCCCGGTCG

50

(2) INFORMATION FOR SEQ ID NO:501:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytomegalovirus DNA
polymerase gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:501:

TAGGCGGGCT GGAAAGATGA TGTATAAATA GAGTCTCCGA CGGGGTTTCG

50

(2) INFORMATION FOR SEQ ID NO:502:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human cytomegalovirus b' 2.2 kb
transcript (start 160513)

463

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:502:

TAGGCGGGCT GGAAGATGA TGTATAAATA GAGTCTGCGA CGGGGTTCGG

50

(2) INFORMATION FOR SEQ ID NO:503:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus 2.7 kb transcript (start 4578)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:503:

GCCCCGCGCTC GGCAGAGCTA CCATATAAAA ACCGAGGGGT TTAGCAGCTT

50

(2) INFORMATION FOR SEQ ID NO:504:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 82K AlkExo (start 27048)

464

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:504:

CAGCACCAGG AGAGGCTTAA GCTCGGGAGG CAGCGCCACC GACGACAGTA

50

(2) INFORMATION FOR SEQ ID NO:505:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene
(start site 106547)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:505:

ATGGGTTGTG GTTATATGCA CTCCTATAA GACTCTCCCC CACCGCCCAC

50

(2) INFORMATION FOR SEQ ID NO:506:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 39k dUTPase
gene (start 106811)

465

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:506:

CGTGTGCGAT AATACACACG CCCATCGAGG CCATGCCTAC ATAAAAGGGC

50

(2) INFORMATION FOR SEQ ID NO:507:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 33K (start site 145165)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:507:

GGCCGGGCGA CCCAGATGTT TACTTAAAAG GCGTGCCGTC CGCCGGCATG

50

(2) INFORMATION FOR SEQ ID NO:508:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 21K (start site 145459)

466

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:508:

CGACGTACGC GATGAGATCA ATAAAAGGGG GCGTGAGGAC CGGGAGGCGG

50

(2) INFORMATION FOR SEQ ID NO:509:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 5 kb transcript (start 86216)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:509:

CCCCACCCCT GCGCGATGTG GATAAAAAGC CAGCGCGGGT GGTTTAGGGT

50

(2) INFORMATION FOR SEQ ID NO:510:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' RNR2 gene (start 89774)

467

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:510:

GGTCCGCCTT CTGGTCCACG CATATAAGCG CGGACTAAAA ACAGGGATGT

50

(2) INFORMATION FOR SEQ ID NO:511:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-2 RNR2 gene
(start site 247)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:511:

TGGTCCGCCT TCTCGTCCAC GCATATAAGC GCGGCCTGAA GACGGGGATG

50

(2) INFORMATION FOR SEQ ID NO:512:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' tk gene
(start site 47911)

468

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:512:

CGCGGTCCCA GGTCCACTTC GCATATTAAG GTGACGCGTG TGGCCTCGAA

50

(2) INFORMATION FOR SEQ ID NO:513:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-2 b' tk gene
(start site 225)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:513:

CGCGGCCCCGA GGTCCACTTC GCATATTAAG GTGACGCGCG TGGCCTCGAA

50

(2) INFORMATION FOR SEQ ID NO:514:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' dbp gene
(start site 62318)

469

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:514:

CGGCACGCCC CCAGGTAAAG TGTACATATA CCAACCGCAT ACCAGACGCA

50

(2) INFORMATION FOR SEQ ID NO:515:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' gB (3.3 Kb) start 56081

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:515:

CCACTCAGCG CGCCGCCTGG CGATATATTC GCGAGCTGAT TATCGCCACC

50

(2) INFORMATION FOR SEQ ID NO:516:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' gD (start 138337)

470

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:516:

CCACTCAGCG CGCCGCCTGG CGATATATTC GCGAGCTGAT TATCGCCACC

50

(2) INFORMATION FOR SEQ ID NO:517:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-2 b' gD (start 5918)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:517:

GGAGTATAAT AGAGTCTTTG TGTTTAAAAC CCGGGCTCGG TGTGGTGTTT

50

(2) INFORMATION FOR SEQ ID NO:518:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' gE (start site 141171)

471

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:518:

GGAGAGGGCC CGCGGCGCAT TTAAGGCGTT GTTGTGTTGA CTTTGCCTCT

50

(2) INFORMATION FOR SEQ ID NO:519:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 ICP gene
(start site 58361)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:519:

AATTATTGCT ACGACATCCG TGCTTGTTTG TGTTCCGTGT CTATATCTCT

50

(2) INFORMATION FOR SEQ ID NO:520:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' tr-4
(start site 136729)

472

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:520:

GGCGGTGCTG TTTGCGGGTT GGCACAAAAA GACCCCGATC CGCGTCTGTG

50

(2) INFORMATION FOR SEQ ID NO:521:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 [U-S] b' tr-9
(start 143245)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:521:

GTGACGTCAA TTGCCCGAGC CGCATAAAGG GCCGGTGGTC CGCCTAGCCG

50

(2) INFORMATION FOR SEQ ID NO:522:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b'g' VP5
(start 40768)

473

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:522:

GGGGTGGGGC GGGGGGGGGG GTATATAAGG CCTGGGATCC CACGTCCCCG

50

(2) INFORMATION FOR SEQ ID NO:523:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b'g' 2.1kb transcript (start 26639)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:523:

CCCGTTAACC CCCCACGTGA TCAGCACGCC ACCGACACCG CAGACGAAAA

50

(2) INFORMATION FOR SEQ ID NO:524:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b'g' a'TIF/VSP (start 105259)

474

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:524:

GGGGCGGCCC GTGCGGGTTG CTTAAATGCG TGGTGGCGAC CACGGGCTGT

50

(2) INFORMATION FOR SEQ ID NO:525:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 b'g' 2.7 kb transcript (start 100998)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:525:

GCCACGCCCA TAAGCTCCTC CCGATAAAAA CCGCCCCGAT GGCCCTGGAC

50

(2) INFORMATION FOR SEQ ID NO:526:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus UL36 gene (start 49862)

475

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:526:

GACGTCAACG CTGATAGTGT CTATAAAGGC CGTGCCGCCG CGCCGTAGTT

50

(2) INFORMATION FOR SEQ ID NO:527:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus g' pp65 gene
(start 121072)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:527:

TCCCGGTTTG GTCGCCTGCC TATGTAAGGC GCGGCCGCA GAGGGCGCGC

50

(2) INFORMATION FOR SEQ ID NO:528:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus g' pp71 gene
(start 119223)

476

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:528:

GTCACCGCTG CTATATTTC GACAGTTGCC GGAACCCTTC CCGACCTCCC

50

(2) INFORMATION FOR SEQ ID NO:529:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human cytomegalovirus g' pp150 gene
(start 43092)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:529:

CGTATCCGCC TCCGCTATTA AACTACCCCC CCTCCCTCTA GGTGGGGCGC

50

(2) INFORMATION FOR SEQ ID NO:530:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 g' 5 Kb
transcript (start 103313)

477

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:530:

TTGTGTCGCA GGGCGGCCCG CGTATAAAGG CGAGAGCGCG GGACCGTTTC

50

(2) INFORMATION FOR SEQ ID NO:531:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 g' gC (start 96170)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:531:

AACCCCGGAT GGGGCCCGGG TATAAATTCC GGAAGGGGAC ACGGGCTACC

50

(2) INFORMATION FOR SEQ ID NO:532:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-2 g' gC (start 670)

478

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:532:

GCGGGGGTGC CGTGGACGGG TATAAAGGCC AGGGGGGCAC GCGGGCCCAT

50

(2) INFORMATION FOR SEQ ID NO:533:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 g' gH gene
(start 46581)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:533:

CGGCAATAAA AAGACAGAAT AAAACGCACG GGTGTTGGGT CGTTTGTTC

50

(2) INFORMATION FOR SEQ ID NO:534:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 g' 42 K
(start site 107130)

479

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:534:

CCGGAGTCCC CGCTAACCTT CGGCATAAAA GCCACCGCGC GCCTGTTGAC

50

(2) INFORMATION FOR SEQ ID NO:535:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 Ori_s ORF
(start site 132287)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:535:

CGGAGGCCCC CGGGGTGCGT CCCCTGTGTT TCGTGGGTGG GGTGGGCGCG

50

(2) INFORMATION FOR SEQ ID NO:536:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-1 18 K (start
site 97951)

480

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:536:

CCCCCCCACC GCTGGGCGCT ATAAAGCCGC CACCCTCTCT TCCCTCAGGT

50

(2) INFORMATION FOR SEQ ID NO:537:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Herpes simplex virus-2 18K (start site 2391)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:537:

CCCCGCCGTC CCCCAGGCGT TATAAGCCGC CGCACTCGCT TTTCCACCG

50

(2) INFORMATION FOR SEQ ID NO:538:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus L1 1Kb gene (start site 103194)

481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:538:

TGGTGCCTTG GCTTTAAAGG GGAGATGTTA GACAGGTAAC TCACTAAACA

50

(2) INFORMATION FOR SEQ ID NO:539:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus R1 145K gene
(start site 1721)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:539:

ACTGTATAAA GGTAAGTATT ATTAAATTTT AGAGACACTA TCACGTGTAA

50

(2) INFORMATION FOR SEQ ID NO:540:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus R1 20K (start
site 9660)

482

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:540:

CTTTTAGCCA TGCCATGCTC TATAAATCAC TTCCCTATCT CAGGTAGGCC

50

(2) INFORMATION FOR SEQ ID NO:541:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [DL/R] (start site 52787)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:541:

ACAGAGACCC CAAAAGAGG ATAAAAGAAG GCGAGCCGGC CCGGCTCGCC

50

(2) INFORMATION FOR SEQ ID NO:542:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site 61372)

483

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:542:

GTGACGGTCA GGCAGCTCCT GTATTAACT TTGCGGACAG AGGCCAGAGC

50

(2) INFORMATION FOR SEQ ID NO:543:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus L2 (start site 57050)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:543:

TAATTACGCT TGTGTACATA TTAAATCCA CACAAGTGGC CAGAGTGGGC

50

(2) INFORMATION FOR SEQ ID NO:544:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus R1 (start site 88539)

484

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:544:

GACAGGGGACG GCGGCGCTAT ATATAAGAGC CCAAGACCCG GCTCTCTTTA

50

(2) INFORMATION FOR SEQ ID NO:545:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site 88897)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:545:

CGGATTAGAT GGGGATATTT AAAAGGGGCA GCAATCTCGG CTGTTTGTAC

50

(2) INFORMATION FOR SEQ ID NO:546:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus L2 (start site 90021)

485

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:546:

ACCCAACAGG TGGTGAAAAT ATAACACAGG TGACACCAGC CTCTATCAGC

50

(2) INFORMATION FOR SEQ ID NO:547:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [BamH1-L] L1
(start site 92157)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:547:

ACCCCCCTTG TACCTATTAA AGAGGATGCT GCCTAGAAAT CGGTGCCGAG

50

(2) INFORMATION FOR SEQ ID NO:548:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [BamH1-L] L3
(start site 88480)

486

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:548:

CGGGTCTTGG GCTCTTATAT ATAGCGCCGC CGTCCCTGTC TGTTAGATCA

50

(2) INFORMATION FOR SEQ ID NO:549:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [BamH1-K] 2.1 Kb
(start site 109939)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:549:

AGACGCCCTC AATCGTATTA AAAGCCGTGT ATTCCCCCGC ACTAAGAAT

50

(2) INFORMATION FOR SEQ ID NO:550:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [BamH1-K] 1.3kb
transcript (start 110632)

487

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:550:

TTGCGACCCC TCTGATATTA AGGTGGTTAT TTTGGGCCAG GACCCCTATC

50

(2) INFORMATION FOR SEQ ID NO:551:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [EcoRI-H] L1
(start site 137680)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:551:

CGGTGCCCCG ACTCAGAATT ATTAACCGG GTGGCAGCTC CTGGCAGTCA

50

(2) INFORMATION FOR SEQ ID NO:552:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [EcoRI-D] L1
(start site 159337)

488

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:552:

AAGGGCAGGG GGTGGGTATT TAAGGATCTA TATGCCCTTC TCTACCTGCA

50

(2) INFORMATION FOR SEQ ID NO:553:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [EcoR1-D] R1
(start 165496)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:553:

AATGGGCGTG GCAGAATAGT ATAAGACCGG AGGCCTGGGT GAGGAGAGTC

50

(2) INFORMATION FOR SEQ ID NO:554:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [EcoR1-D] L2
(start 167495)

489

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:554:

TCTTTCCTTG TCCTTACTGT ATAAAAGTCC ACGAAAACAG CTGTGCCTCA

50

(2) INFORMATION FOR SEQ ID NO:555:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Epstein Barr Virus [EcoR1-D] L1A
start 169165

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:555:

ACTGATGAGT AAGTATTACA CCCTTTGCCC CACACCCCCT TTCCCTTACT

50

(2) INFORMATION FOR SEQ ID NO:556:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Epstein Barr Virus [EBNA] E1 (start
site 11333)

490

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:556:

AGGGGGGGAC TAAGGTCCCA CTACAAAAC TCTGTGTTCT GCTGCAAATT

50

(2) INFORMATION FOR SEQ ID NO:557:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [EBNA] E2 (start site 14399)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:557:

GGTATAAAGT GGTCTGCAG CTATTTCTGG TCGCATCAGA GCGCCAGGAG

50

(2) INFORMATION FOR SEQ ID NO:558:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [EcoRI-D] L1 (start site 169514)

491

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:558:

CTCTGACGTA GCCGCCCTAC ATAAGCCTCT CACACTGCTC TGCCCCCTTC

50

(2) INFORMATION FOR SEQ ID NO:559:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type 2 E1a (start 498)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:559:

GTCAGCTGAC GCGCAGTGTA TTTATACCCG GTGAGTTCCT CAAGAGGCCA

50

(2) INFORMATION FOR SEQ ID NO:560:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-5 E1a (start 499)

492

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:560:

GTCAGCTGAC GTGTAGTGTA TTTATACCCG GTGAGTTCCT CAAGAGGCCA

50 -

(2) INFORMATION FOR SEQ ID NO:561:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-7 E1a (start site 512)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:561:

TCAGCTGATC GCTAGGGTAT TTAAACCTGA CGAGTTCCT CAAGAGGCCA

50

(2) INFORMATION FOR SEQ ID NO:562:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-12 E1a (start site 306)

493

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:562:

AAATTGATGA CGGCAATTTT ATTATAGGCG CGGAATATTT ACCGAGGGCA

50

(2) INFORMATION FOR SEQ ID NO:563:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-12 E1a (start site 445)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:563:

GTCAGCTGAT CGTTGGGTA TTTAATGCCG CCGTGTCGT CAAGAGGCCA

50

(2) INFORMATION FOR SEQ ID NO:564:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Simian Adenovirus SA7 E1a (start site 440)

494

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:564:

TTATTGTCTA GGTGAGGGTA TTAAACCGG CTCAGACCGT CAAGAGGCCA

50

(2) INFORMATION FOR SEQ ID NO:565:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-2 E1b (start 1700)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:565:

GGGGCGGGGC TTAAAGGGTA TATAATGCGC CGTGGGCTAA TCTTGTTAC

50

(2) INFORMATION FOR SEQ ID NO:566:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-5 E1b (start site
1703)

495

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:566:

GGGGCGGGGC TTAAAGGGTA TATAATGCGC CGTGGGCTAA TCTTGTTAC

50

(2) INFORMATION FOR SEQ ID NO:567:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-7 E1b (start site 1577)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:567:

TTCTTGGGTG GGGTCTTGGA TATATAAGTA GGAGCAGATC TGTGTGGTTA

50

(2) INFORMATION FOR SEQ ID NO:568:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-12 E1b (start site 1527)

496

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:568:

TGGGCGTGGT TAAACAGGGA TATAAAGCTG GGTGGTGTG GCTTTGAATA

50

(2) INFORMATION FOR SEQ ID NO:569:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-2 EII (start site 27092)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:569:

GAAAGGGCGC GAAACTAGTC CTTAAGAGTC AGCGCGCAGT ATTTGCTGAA

50

(2) INFORMATION FOR SEQ ID NO:570:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-2 EIII (start site 27610)

497

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:570:

TGCGGTCGCC CGGGCAGGGT ATAAC TCACC TGAAAATCAG AGGGCGAGGT

50

(2) INFORMATION FOR SEQ ID NO:571:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-5 EIII (start site 239)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:571:

TGCGGTCGCC CGGGCAGGGT ATAAC TCACC TGACTCTTGG AGGGCGAGGT

50

(2) INFORMATION FOR SEQ ID NO:572:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-2 EIV (start site 35611)

498

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:572:

TTACGTCATT TTTAGTCCT ATATATACTC GCTCTGTACT TGGCCCTTTT

50

(2) INFORMATION FOR SEQ ID NO:573:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-2 IVa2 (start site 5827)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:573:

CCCTCCCACT TAGCCTCCTT CGTGCTGGCC TGGACGCGAG CCTTCGTCTC

50

(2) INFORMATION FOR SEQ ID NO:574:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-5 IVa2 (start site 5837)

499

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:574:

CCCTCCCACT TAGCCTCCTT CGTGCTGGCC TGGACGCGAG CCTTTGTCTC

50

(2) INFORMATION FOR SEQ ID NO:575:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-7 IVa2 (start site 5692)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:575:

CCCTCCCACG TGGCCTCCTT TGTGCTGGCC TGGACACGCG CTTTTGTATC

50

(2) INFORMATION FOR SEQ ID NO:576:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-2 IX (start site 3575)

500

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:576:

GCTTAAGGGT GGGAAAGAAT ATATAAGGTG GGGGTCTCAT GTAGTTTGT

50

(2) INFORMATION FOR SEQ ID NO:577:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-5 IX (start site 3581)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:577:

GCTTAAGGGT GGGAAAGAAT ATATAAGGTG GGGGTCTTAT GTAGTTTGT

50

(2) INFORMATION FOR SEQ ID NO:578:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-7 IX (start site 3460)

501

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:578:

ATGGGGACTT TCAGGTTGGT AAGGTGGACA AATTGGGTAA ATTTTGTTAA

50

(2) INFORMATION FOR SEQ ID NO:579:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-2 major late
(start site 6039)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:579:

GTGTTCTGA AGGGGGGCTA TAAAGGGGG TGGGGGCGCG TCGTCCTCA

50

(2) INFORMATION FOR SEQ ID NO:580:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Adenovirus type-5 major late (start
site 6049)

502

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:580:

GTGTTCTGA AGGGGGGCTA TAAAAGGGGG TGGGGGCGCG TTCGTCCTCA

50

(2) INFORMATION FOR SEQ ID NO:581:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-7 major late (start site 5904)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:581:

GGGTCCCCGC CGGGGGGTA TAAAAGGGGG CGGACCTCTG TTCGTCCTCA

50

(2) INFORMATION FOR SEQ ID NO:582:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-12 major late (start site 972)

503

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:582:

AATTTTCTGG TGGTGGGCTA TAAAAAGGGG CGGGTCCTTG GTCTTCATCG

50

(2) INFORMATION FOR SEQ ID NO:583:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Adenovirus type-2 LIIa (start site 25954)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:583:

GGCGTGGTAG TCCTCAGSTA CAAATTGCG AAGGTAAGCC GACGTCCACA

50

(2) INFORMATION FOR SEQ ID NO:584:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human papilloma virus type 18 E6 gene (start site 30)

504

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:584:

CAGCACATAC TATACTTTTC ATTAATACTT TTAACAATTG TAGTATATAA

50

(2) INFORMATION FOR SEQ ID NO:585:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human papilloma virus type-16 E6/E7
(start site 97)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:585:

GAACCGAAAC CGGTTAGTAT AAAAGCAGAC ATTTTATGCA CCAAAAGAGA

50

(2) INFORMATION FOR SEQ ID NO:586:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Human papilloma virus type-18 E6
(start site 105)

505

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:586:

GGACCGAAAA CGGTGTATAT AAAAGATGTG AGAAACACAC CACAATACTA

50

(2) INFORMATION FOR SEQ ID NO:587:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Parvovirus h-1 H-1[+.04] (start site 209)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:587:

AGTGGGCGTG GCTAACTGTA TATAAGCAGT CACTCTGGTC GGTTACTCAC

50

(2) INFORMATION FOR SEQ ID NO:588:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Parvovirus h-1 H-1 [+.40] (start site 2010)

506

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:588:

GCCGAAGCTA GACACTCCTA TAAATTCGCT AGGTTCAATG CGCTCACCAT

50

(2) INFORMATION FOR SEQ ID NO:589:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Human parvovirus B19-Au B19 [0.06]
(start site 347)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:589:

GAGCGTAGGC GGGGACTACA GTATATATAG CACGGTACTG CCGCAGCTCT

50

(2) INFORMATION FOR SEQ ID NO:590:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Simian virus 40 T/t late (start
site 31)

507

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:590:

CCGCCCCTAA CTCCGCCAG TTCCGCCAT TCTCCGCCCC ATGGCTGACT

50

(2) INFORMATION FOR SEQ ID NO:591:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Simian virus 40 T/t early P2 (start site 5233)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:591:

TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCGGCCT

50

(2) INFORMATION FOR SEQ ID NO:592:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: BK virus T/t early (start site 99)

508

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:592:

CCTGTGGCCT TTTTTTTAT AATATATAAG AGGCCGAGGC CGCCTCTGCC

50

(2) INFORMATION FOR SEQ ID NO:593:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Polyoma virus T/t E (start site 156)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:593:

GGCCACCCAA ATTGATATAA TTAAGCCCCA ACCGCCTCTT CCCGCCTCAT

50

(2) INFORMATION FOR SEQ ID NO:594:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Simian virus 40 late (start site 325)

509

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:594:

GTTCTTTCCG CCTCAGAAGG TACCTAACCA AGTTCCTCTT TCAGAGGTTA

50

(2) INFORMATION FOR SEQ ID NO:595:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Hepatitis B virus subtype adr4 3.6kb
P1 (start 1659)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:595:

AGTTGGGGGA GGAGATTAGG TTAAAGGTCT TTGTACTAGG AGGCTGTAGG

50

(2) INFORMATION FOR SEQ ID NO:596:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Hepatitis B virus subtype adr4 3.6 kb
P2 (start 1690)

510

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:596:

TGTACTAGGA GGCTGTAGGC ATAAATTGGT CTGTTACCA GCACCATGCA

50

(2) INFORMATION FOR SEQ ID NO:597:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Hepatitis B virus subtype adr4 2.2
kb P1 (start 3061)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:597:

ATCGGCACTC AGGAAGACAG CCTACTCCCA TCTCTCCACC TCTAAGAGAC

50

(2) INFORMATION FOR SEQ ID NO:598:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Hepatitis B virus subtype adr4 2.2
kb P2 (start 3092)

511

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:598:

CTCTCCACCT CTAAGAGACA GTCATCCTCA GGCCATGCAG TGGAAGTCCA

50

(2) INFORMATION FOR SEQ ID NO:599:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Epstein Barr virus [BamH1-F] R1
(start 58862)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:599:

TATTTTGTGAA AAGGGATATT ATAAACAGG TCATTGCTCG GATTGTGGCA

50

(2) INFORMATION FOR SEQ ID NO:600:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Promoter Sequenc of IL-13

512

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:600:

GGTGTGAGGC GTCACCACTT GGGCCTATAA AAGCTGCCAC AAGACGCCAA GGCCAC

56

(2) INFORMATION FOR SEQ ID NO:601:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: UL9 BINDING SITE, HSV oris, higher affinity

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:601:

CGTTGCGCACT T

11

(2) INFORMATION FOR SEQ ID NO:602:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

513

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 BINDING SITE, HSV oris, lower
affinity

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:602:

TGCTCGCACT T

11

(2) INFORMATION FOR SEQ ID NO:603:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL921 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:603:

GCGCGCGCGC GTTCGCACTT CCGCCGCCGG

30

(2) INFORMATION FOR SEQ ID NO:604:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

514

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL922 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:604:

GGCGCCGGCC GTTCGCACTT CGCGCGCGCG

30

(2) INFORMATION FOR SEQ ID NO:605:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 CCCG TEST SEQ. / UL9 ASSAY
SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:605:

GGCCCCCCCC GTTCGCACTT CCCGCCCCGG

30

(2) INFORMATION FOR SEQ ID NO:606:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

515

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 GGGC TEST SEQ. / UL9 ASSAY
SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:606:

GGCGGGCGCC GTTCGCACTT GGGCGGGCGG

30

(2) INFORMATION FOR SEQ ID NO:607:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 ATAT TEST SEQ. / UL9 ASSAY
SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:607:

GGATATATAC GTTCGCACTT TAATTATTGG

30

(2) INFORMATION FOR SEQ ID NO:608:

516

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: UL9 polyA TEST SEQ. / UL9 ASSAY
SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:608:

GGAAAAAAAC GTTCGCACTT AAAAAAAGG

30

(2) INFORMATION FOR SEQ ID NO:609:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: UL9 polyT TEST SEQ. / UL9 ASSAY
SEQ.

517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:609:

GGTTTTTTTC GTTCGCACTT TTTTTTTTGG

30

(2) INFORMATION FOR SEQ ID NO:610:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 GCAC TEST SEQ. / UL9 ASSAY
SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:610:

GGACGCACGC GTTCGCACTT GCAGCAGCGG

30

(2) INFORMATION FOR SEQ ID NO:611:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

518

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 ATori-1 Test sequence / UL9
ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:611:

GCGTATATAT CGTTCGCACT TCGTCCCAAT

30

(2) INFORMATION FOR SEQ ID NO:612:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: orieCO2 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:612:

GGCGAATTCG ACGTTCGCAC TTCGTCCCAA T

31

(2) INFORMATION FOR SEQ ID NO:613:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

519

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: oriECO3 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:613:

GGCGAATTCG ATCGTTCGCA CTTCGTCCCA AT

32

(2) INFORMATION FOR SEQ ID NO:614:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: WILD TYPE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:614:

AAGTGAGAAT TCGAAGCGTT CGCACTTCGT CCCAAT

36

(2) INFORMATION FOR SEQ ID NO:615:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

520

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: TRUNCATED UL9 BINDING SITE, COMPARE
SEQ ID NO:601

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:615:

TTCGCACTT

9

(2) INFORMATION FOR SEQ ID NO:616:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HSVB1/4, SEQUENCE OF COMPETITOR DNA
MOLECULE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:616:

GGTCGTTTCGC ACTTCGC

17

(2) INFORMATION FOR SEQ ID NO:617:

521

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Figure 14B, top strand of an exemplary target sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:617:

GCGTANNNNN CGTTCGCACT TNNNNCTTCG TCCCAAT

37

(2) INFORMATION FOR SEQ ID NO:618:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: HSV primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:618:

ATTGGGACGA AG

12

(2) INFORMATION FOR SEQ ID NO:619:

522

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: a sample distamycin target sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:619:

TTCCTCCTTT C

11

(2) INFORMATION FOR SEQ ID NO:620:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: a distamycin target sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:620:

TTCCNNNTTT C

11

(2) INFORMATION FOR SEQ ID NO:621:

523

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Figure 27A, test oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:621:

GCGTANNNNN CGTTCGCACT TNNNNCTTCG TCCCAAT

37

(2) INFORMATION FOR SEQ ID NO:622:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Figure 27B, oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:622:

GCGTANNNNN CGTTCGCACT TNNNNCTTCG TCCCAAT

37

(2) INFORMATION FOR SEQ ID NO:623:

(i) SEQUENCE CHARACTERISTICS:

524

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Figure 27C, oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:623:

GCGTANNNNN TTCACGCTTG CNNNNCTTCG TCCCAAT

37

(2) INFORMATION FOR SEQ ID NO:624:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Figure 27D, oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:624:

GCGTANNNNN TTCACGCTTG CNNNNCTTCG TCCCAAT

37

(2) INFORMATION FOR SEQ ID NO:625:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 base pairs

525

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: -35 region consensus sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:625:

TTGACA

6

(2) INFORMATION FOR SEQ ID NO:626:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: -10 region consensus sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:626:

TATAAT

6

(2) INFORMATION FOR SEQ ID NO:627:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 242 base pairs
- (B) TYPE: nucleic acid

526

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HIV-1, LTR sequence, Figure 28

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:627:

GTTAGAGTGG AGGTTTGACA GCCGCCTAGC ATTCATCAC ATGGCCCGAG AGCTGCATCC	60
GGAGTACTTC AAGAACTGCT GACATCGAGC TTGCTACAAG GGACTTTCCG CTGGGGACTT	120
TCCAGGGAGG CGTGGCCTGG GCGGGACTGG GGAGTGGCGA GCCCTCAGAT CCTGCATATA	180
AGCAGCTGCT TTTTGCTGT ACTGGGTCTC TCTGGTTAGA CCAGATCTGA GCCTGGGAGC	240
TC	242

(2) INFORMATION FOR SEQ ID NO:628:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: a TFIID binding site

527

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:628:

CCTGCATA

8

(2) INFORMATION FOR SEQ ID NO:629:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: a TFIID binding site

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:629:

AAGCAGCT

8

(2) INFORMATION FOR SEQ ID NO:630:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 29A

528

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:630:

GCAGAATTCT GCAG

14

(2) INFORMATION FOR SEQ ID NO:631:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 29A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:631:

GCAGAATTCT GCAGCGTTTCG CACTTTCTAG AGCTCAGG

38

(2) INFORMATION FOR SEQ ID NO:632:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 29A

529

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:632:

AGATCTCGAG TCC

13

(2) INFORMATION FOR SEQ ID NO:633:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 29B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:633:

GCAGAATTCT GCAGNNNNCG TTCGCACTTT CTAGAGCTCA GG

42

(2) INFORMATION FOR SEQ ID NO:634:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 29C

530

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:634:

GCAGAATTCT GCAGNNNNNN NNCGTTCGCA CTTTCTAGAG CTCAGG

46

(2) INFORMATION FOR SEQ ID NO:635:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 29D

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:635:

GCAGAATTCT GCAGCGTTCG CACTTNNNNN NNNTCTAGAG CTCAGG

46

(2) INFORMATION FOR SEQ ID NO:636:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 30

531

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:636:

CGTGAATTCT GCAG

14

(2) INFORMATION FOR SEQ ID NO:637:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:637:

CGTGAATTCT GCAGATG

17

(2) INFORMATION FOR SEQ ID NO:638:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 30

532

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:638:

CGTGAATTCT GCAGATGAGG TACCNNNNNN CGTTCGCACT TTCTAGAGCT CTCC

54

(2) INFORMATION FOR SEQ ID NO:639:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:639:

GTGAAAGATC TCGAGAGG

18

(2) INFORMATION FOR SEQ ID NO:640:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Oligonucleotide, Figure 30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:640:

AAGATCTCGA GAGG

14

533

(2) INFORMATION FOR SEQ ID NO:641:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: UL9 BINDING SITE, HSV oris

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:641:

CGTTCTCACT T

11

534

IT IS CLAIMED:

1. A method of constructing a DNA-binding agent capable of sequence-specific binding to a duplex DNA target region, comprising:
 - identifying in the duplex DNA, a target region containing a series of at least two non-overlapping base-pair sequences of four base-pairs each, where the four base-pair sequences are adjacent and each sequence is characterized by sequence-preferential binding to a small molecule, and
 - coupling the small molecules to form a DNA-binding agent capable of sequence-specific binding to said target region.
2. The method of claim 1, where the duplex binding small molecules are identified as molecules capable of binding to a selected test sequence in a duplex DNA by:
 - (i) adding a molecule to be screened to a test system composed of (a) a duplex DNA test oligonucleotide having a screening sequence adjacent a selected test sequence, where a DNA binding protein is effective to bind to said screening sequence with a binding affinity that is substantially independent of such test sequence, but where DNA protein binding to the screening sequence is sensitive to binding of test molecules to such test sequence, and (b) said DNA-binding protein,
 - (ii) incubating the molecule in the test system for a period sufficient to permit binding of the molecule being tested to the test sequence in the duplex DNA, and

535

(iii) comparing the amount of binding protein bound to the duplex DNA test oligonucleotide before and after said adding.

5 3. The method of claim 2, where the screening sequence is from the HSV origin of replication and the binding protein is UL9.

10 4. The method of claim 3, wherein the DNA screening sequence is selected from the group consisting of SEQ ID NO:601, SEQ ID NO:602, SEQ ID NO:615, and SEQ ID NO:641.

15 5. The method of claim 2, where said comparing is accomplished using either a gel band-shift assay or a filter-binding assay.

20 6. The method of claim 1, where the subunits are the same and the DNA-binding agent is a homopolymer.

 7. The method of claim 1, where the subunits are different and the DNA-binding agent is a heteropolymer.

25 8. The method of claim 1, where the four base pair sequences are separated by at least 1-6 base-pairs.

30 9. The method of claim 1, where said DNA-binding small molecules are coupled to each other using a spacer molecule.

 10. The method of claim 1, where the two sequences are selected from the group of sequences consisting of TTTC, TTTG, TTAC, TTAG, TTGC, TTGG, TTCC, TTCG,

536

TATC, TATG, TAAC, TAAG, TAGC, TAGG, TACC, TAGC sequences.

11. The method of claim 10, where the duplex DNA-binding small molecule is distamycin.

12. A method of blocking transcription activity from a duplex DNA template, comprising
identifying in the duplex DNA a binding site
10 for a transcription factor and, adjacent the binding site, a target region having series of at least two non-overlapping base-pair sequences of four base-pairs each, where the four base-pair sequences are adjacent and each sequence is characterized by sequence-preferential binding to a small molecule, and
15 contacting the duplex DNA with a binding agent composed of the small molecules coupled to form a DNA-binding agent capable of sequence-specific binding to said target region.

20

13. The method of claim 12, where the target region is selected from DNA sequences adjacent a binding site for a eucaryotic transcription factor.

25 14. The method of claim 13, where the transcription factor is TFIID.

15. The method of claim 12, where the target region is selected from DNA sequences adjacent a
30 binding site for a procaryotic transcription factor.

16. The method of claim 15, where the transcription factor is a sigma factor.

17. A DNA-binding agent capable of binding with base-sequence specificity to a target region in duplex DNA, where the target region contains at least two adjacent four base-pair sequences, comprising

5 at least two subunits, where each subunit is a small molecule and has a sequence-preferential binding affinity for a sequence of four base-pairs in the target region, and

 where the subunits are coupled to form a DNA-binding agent capable of sequence-specific binding to said target region.

18. A method of constructing a binding agent capable of sequence-specific binding to a duplex DNA target region, comprising:

15 identifying in the duplex DNA, a target region containing (i) a series of at least two adjacent non-overlapping base-pair sequences of four base-pairs each, where each four base-pair sequence is characterized by sequence-preferential binding to a small molecule, and (ii) adjacent to (i) a DNA duplex region capable of forming a triplex with a third-strand oligonucleotide,

20 coupling the small molecules to form a DNA-binding agent capable of sequence-specific binding to said target region, and

 attaching the DNA-binding agent to a third strand oligonucleotide.

30 19. The method of claim 18, where binding of the DNA-binding agent to duplex DNA causes a shift from B form to A form DNA.

20. A DNA-binding agent capable of binding with base-sequence specificity to a duplex DNA target region

538

containing two sites, a first site having at least two adjacent four base pair sequences, and a second site capable of forming a triplex with a third-strand oligonucleotide, said DNA-binding agent comprising

5 at least two subunits, where each subunit is a small molecule and has a sequence-preferential binding affinity for a sequence of four base-pairs in the target region, where the subunits are coupled to form a DNA-binding agent capable of sequence-specific
10 binding to said first site, and

 a third strand capable of forming a triplex with the second site,

 where the third strand is attached to the DNA-binding agent.

15

21. A method of ordering the sequence binding preferences of a DNA-binding molecule, comprising

(i) adding a molecule to be screened to a test system composed of (a) a duplex DNA test oligonucleotide having a screening sequence adjacent a
20 selected test sequence, where a DNA binding protein is effective to bind to said screening sequence with a binding affinity that is substantially independent of such test sequence, but where DNA protein binding to
25 the screening sequence is sensitive to binding of test molecules to such test sequence, and (b) said DNA-binding protein,

(ii) incubating the molecule in the test system for a period sufficient to permit binding of the
30 molecule being tested to the test sequence in the duplex DNA,

(iii) comparing the amount of binding protein bound to the duplex DNA test oligonucleotide before and after said adding,

539

(iv) repeating steps (i) and (iii) using duplex DNA test oligonucleotides containing all test sequences of interest, and

5 (v) ordering the relative amounts of protein to each duplex DNA test oligonucleotide bound in the presence of the molecule for each test sequence.

22. The method of claim 21, where the test sequences are selected from the group of 256 possible
10 four base sequences composed of A, G, C and T.

23. A method for altering the binding characteristics of a DNA-binding protein to a duplex DNA, comprising

15 identifying in the duplex DNA (i) a binding site for the DNA-binding protein, where said site comprises a series of contiguous paired nucleotides, and (ii) a target region adjacent the binding site,

selecting a small molecule characterized by
20 sequence-preferential binding to the target region, where, when the small molecule is bound to the target region, the small molecule is adjacent to the site for the DNA-binding protein or overlapping the site for the DNA-binding protein by at least one nucleotide pair,
25 and

contacting the duplex DNA with the small molecule at a concentration effective to alter binding of the DNA-binding protein to its binding site.

30 24. The method of claim 23, where contacting the duplex DNA with the small molecule inhibits the binding of the DNA-binding protein to its binding site.

540

25. The method of claim 23, where contacting the duplex DNA with the small molecule enhances the binding of the DNA-binding protein to its binding site.

5 26. The method of claim 23, where the DNA binding protein is a eucaryotic general transcription factor and the target region is selected from DNA sequences adjacent the binding site for the eucaryotic transcription factor.

10

27. The method of claim 26, where the transcription factor is TFIID.

15 28. The method of claim 27, where the region is selected from the group of DNA sequences consisting of SEQ ID NO:1 to SEQ ID NO:600.

20 29. The method of claim 23, where the DNA binding protein is a eucaryotic general transcription factor and the small molecule binds, in addition to the target region, 1 to three nucleotide pairs of the DNA-binding protein's binding site.

25 30. The method of claim 29, where the eucaryotic general transcription factor is TFIID, and the small molecule binds to (i) the target region, and (ii) up to two nucleotides of the binding site for the eucaryotic transcription factor, where the nucleotides are contiguous to the target region.

30

31. The method of claim 23, where the DNA binding protein is a DNA replication factor.

541

32. A method of identifying test sequences in duplex DNA to which binding of a test molecule is most preferred, comprising

(i) constructing a mixture of duplex DNA test
5 oligonucleotides, where each oligonucleotide has (a) a screening sequence adjacent (b) a test sequence, where a DNA binding protein is effective to bind to said screening sequence with a binding affinity that is substantially independent of such test sequence, but
10 where DNA protein binding to the screening sequence is sensitive to binding of test molecules to such test sequence, and (c) where test oligonucleotides of the mixture contain different test sequences,

(ii) adding a test molecule to be screened to a
15 test reaction composed of (a) said DNA binding protein, and (b) said duplex DNA test oligonucleotide mixture,

(iii) incubating the molecule in the test reaction for a period sufficient to permit binding of the molecule being tested to test sequences in the duplex
20 DNA,

(iv) separating test oligonucleotides from test oligonucleotides bound to binding protein,

(v) amplifying the separated test oligonucleotides,

25 (vi) repeating steps (ii) to (v),

(vii) isolating the amplified test oligonucleotides,

(viii) sequencing the isolated test oligonucleotides.

30

33. The method of claim 32, where said test sequences are selected from the group of 256 possible four base sequences composed of A, G, C and T.

542

34. The method of claim 32, where said constructing includes selecting test sequences from the sequences presented as SEQ ID NO:1 to SEQ ID NO:600.

5 35. The method of claim 32, where in constructing the mixture of test oligonucleotides, said adjacent screening and test sequences are flanked by primer sequences.

10 36. The method of claim 35, wherein said amplifying is carried out by successively repeating the steps of (a) denaturing the duplex test oligonucleotides to produce single-strand fragments, (b) hybridizing the single strands with primers, complementary to the
15 primer sequences in the oligonucleotides, to form strand/primer complexes, (c) generating double-strand fragments from the strand/primer complexes in the presence of DNA polymerase and all four deoxyribonucleotides, and (d) repeating steps (a) to (c) until
20 a desired degree of amplification has been achieved.

37. The method of claim 32, wherein said amplifying is carried out by cloning the separated test oligonucleotides into a vector, passaging vectors carrying
25 the test oligonucleotides in appropriate host cells, culturing the host, isolating the vectors, and obtaining the test oligonucleotides from the vectors.

38. The method of claim 32, where said isolating
30 is accomplished by cloning the amplified test oligonucleotides into a cloning vector.

39. The method of claim 32, where said separating
is accomplished by passing the test reaction through a
35 filter, where said filter is capable of capturing

543

DNA:protein complexes but not DNA that is free of protein.

40. The method of claim 32, where the DNA
5 screening sequence is from the HSV origin of replication and the binding protein is UL9.

41. A method of screening for molecules capable
of binding to a selected test sequence in a duplex DNA,
10 comprising

(i) constructing a duplex DNA test oligonucleotide
having a screening sequence adjacent a selected test
sequence, where a DNA binding protein is effective to
bind to said screening sequence with a binding affinity
15 that is substantially independent of such test sequence, but where DNA protein binding to the screening sequence is sensitive to binding of test molecules to such test sequence,

(ii) adding a test molecule to be screened to a
20 test system composed of (a) said DNA binding protein, and (b) said duplex DNA test oligonucleotide having said screening and test sequences adjacent one another,

(iii) incubating the molecule in the test system
for a period sufficient to permit binding of the
25 molecule being tested to the test sequence in the duplex DNA, and

(iv) comparing the amount of binding protein bound to the duplex DNA before and after said adding.

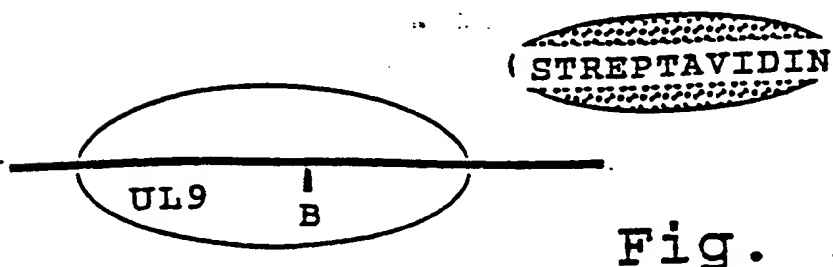
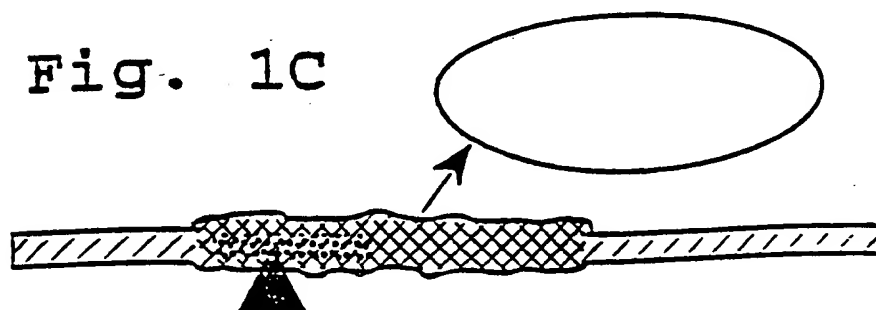
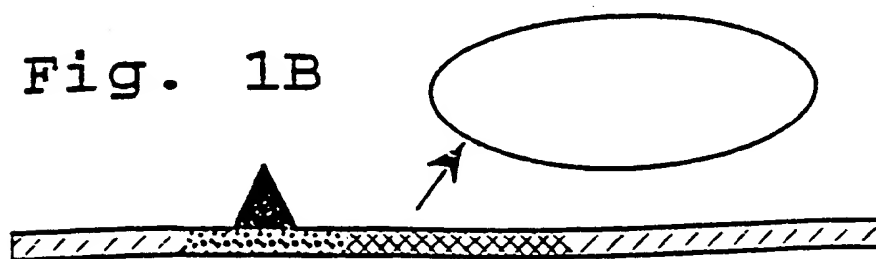
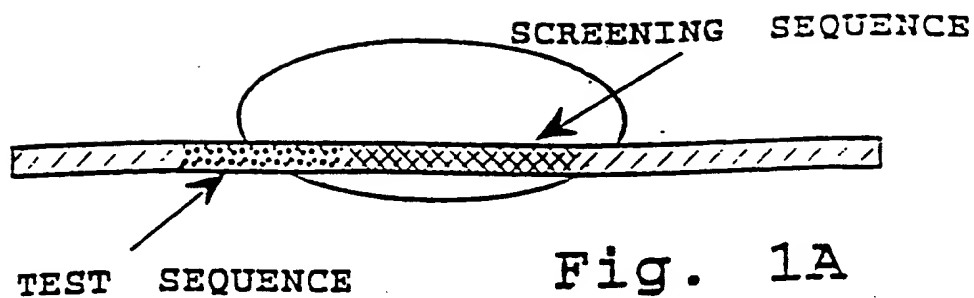
30 42. The method of claim 41, where said test sequence is selected from the group consisting of SEQ ID NO:1 to SEQ ID NO:600.

544

43. The method of claim 41, where the DNA screening sequence is from the HSV origin of replication and the binding protein is UL9.

- 5 44. The method of claim 43, wherein the DNA screening sequence is selected from the group consisting of SEQ ID NO:601, SEQ ID NO:602, SEQ ID NO:615, AND SEQ ID NO:641.

1/39



2 / 39

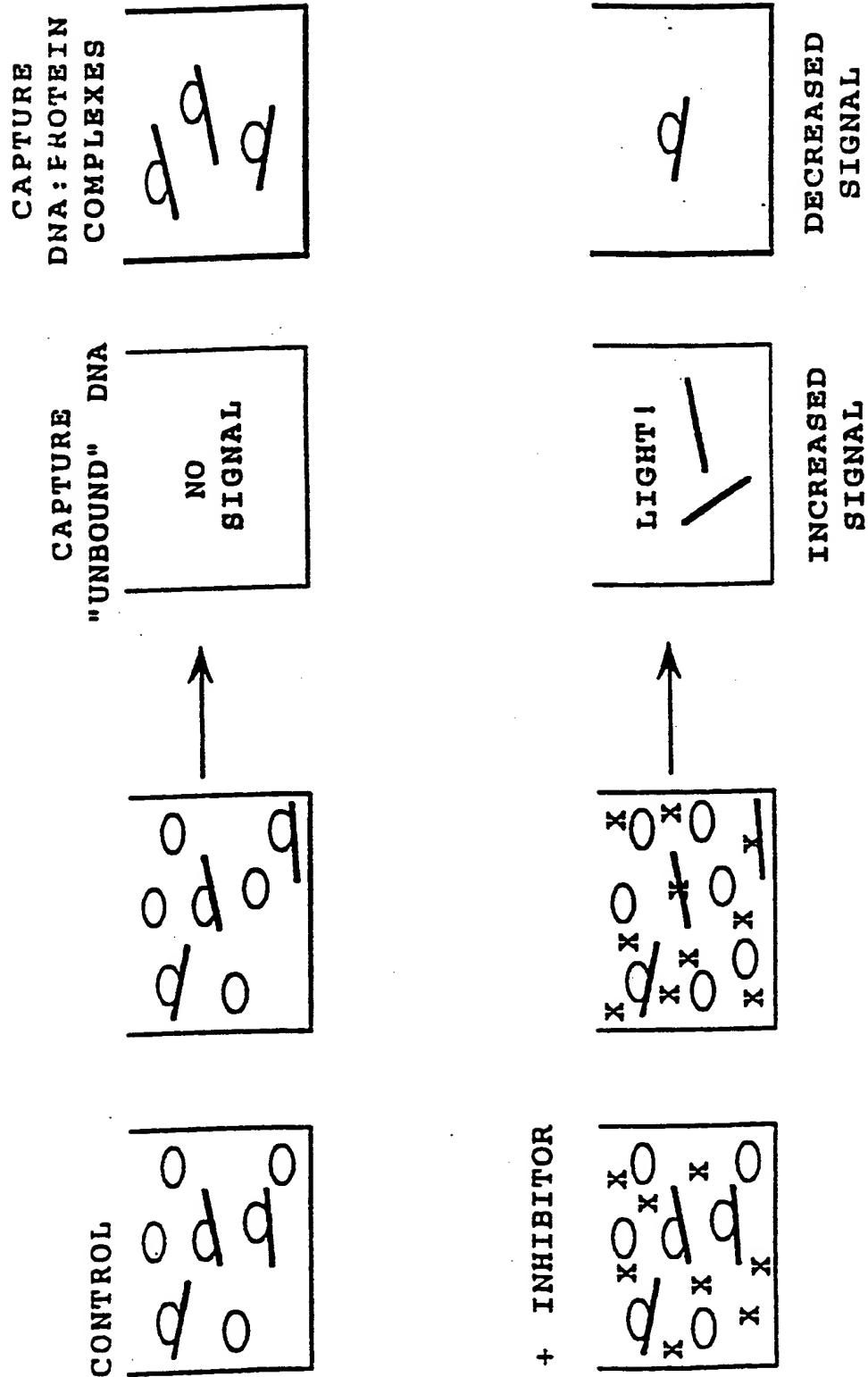


Fig. 2

10
 12345678911
 5' AAGTGAGAATTTCGAAGCGTTCGCACTTCGTCCCAAT 3'
 GAAGCAGGGTTA 5'

+ BIOTIN-11-dUTP
+ KLENOW ENZYME

5' AAGTGAGAATTTCGAAGCGTTCGCACTTCGTCCCAAT 3'
UGAAGCAGGGTTA 5'

PURIFY, THEN ADD
dNTPs + KLENOW

5' AAGTGAGAAATTCGAAGCGTTTCGCACTTCGTCCTCCAAT 3'
3' TTCCTCTTAAGCTTCGCAAGCGUGAAGCAGGGTTA 5'

PURIFY, THEN ADD
DIG-11-OUTP +
TERMINAL TRANSFERASE

5' AAGTGAGAATTCGAAGCGTTCGCACCTTCGTCCCAATUUU 3' DDD
3' UUUTTCACCTCTTAAGCTTCGCAAGCGUGAAGCAGGGTTA 5' B

PURIFY

A diagram of a beam with a central hatched section. The beam is represented by two parallel lines. The left end is labeled 'DDD' and the right end is labeled 'DDD'. A downward arrow points to the center of the beam. Below the beam, a vertical line segment is labeled 'B'.

Fig. 4

4 / 39

	Screening
Test Sequence:	Sequence: Test Sequence
UL9Z1	5'- <u>GCGCGCGCGCGTTCGCACTTCGCGCGCGCGG</u> -3' Z-DNA
UL9Z2	5'-GGCGCGCGCGCGTTCGCACTTCGCGCGCGCGG-3' Z-DNA
UL9 CCGG	5'-GGCCCCGCCCCGTTTCGCACTTCGCGCGCGCGG-3'
UL9 GGGC	5'-GGCGGGGCGCGCGTTCGCACTTCGCGCGCGCGG-3'
UL9 ATAT	5'-GGATATATACGTTTCGCACTTTAATTATTGG-3'
UL9 polyA	5'-GGAAAAAAAAACGTTTCGCACTTAAAAAAAAAGG-3'
UL9 polyT	5'-GGTTTTTTTTTCGTTTCGCACTTTTTTTTTTTGG-3'
UL9 GCAC	5'-GGACGCACGCGTTTCGCACTTGCAGCAGCGG-3'
ATori-1	5'-GCGTATATATACGTTTCGCACTTCGTCCCAAT-3'
oriEco2	5'-GGCGAATTCGACGTTTCGCACTTCGTCCCAAT-3'
oriEco3	5'-GGCGAATTCGATCGTTTCGCACTTCGTCCCAAT-3'

Fig. 5

5 / 39

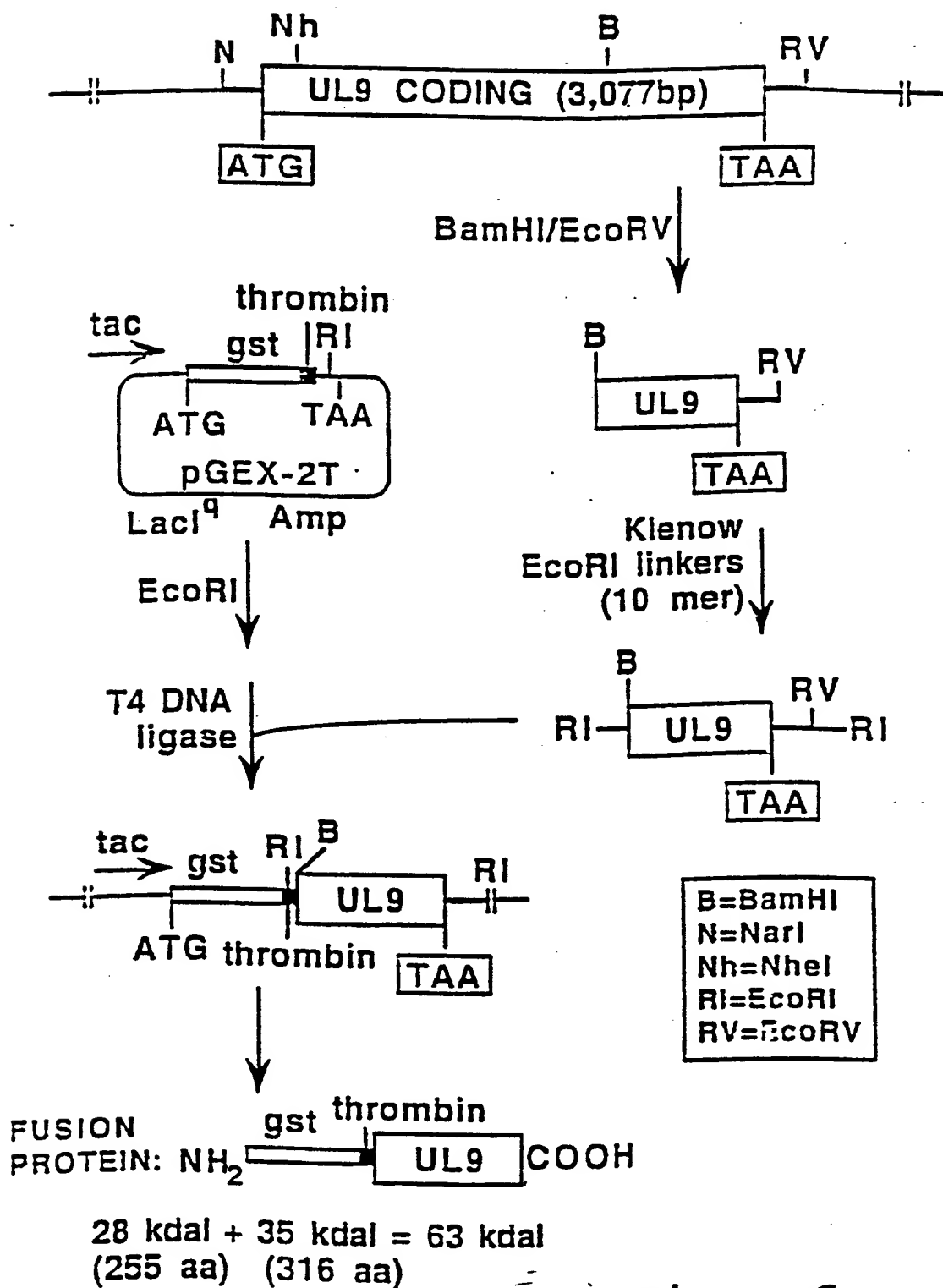


Fig. 6

6 / 39

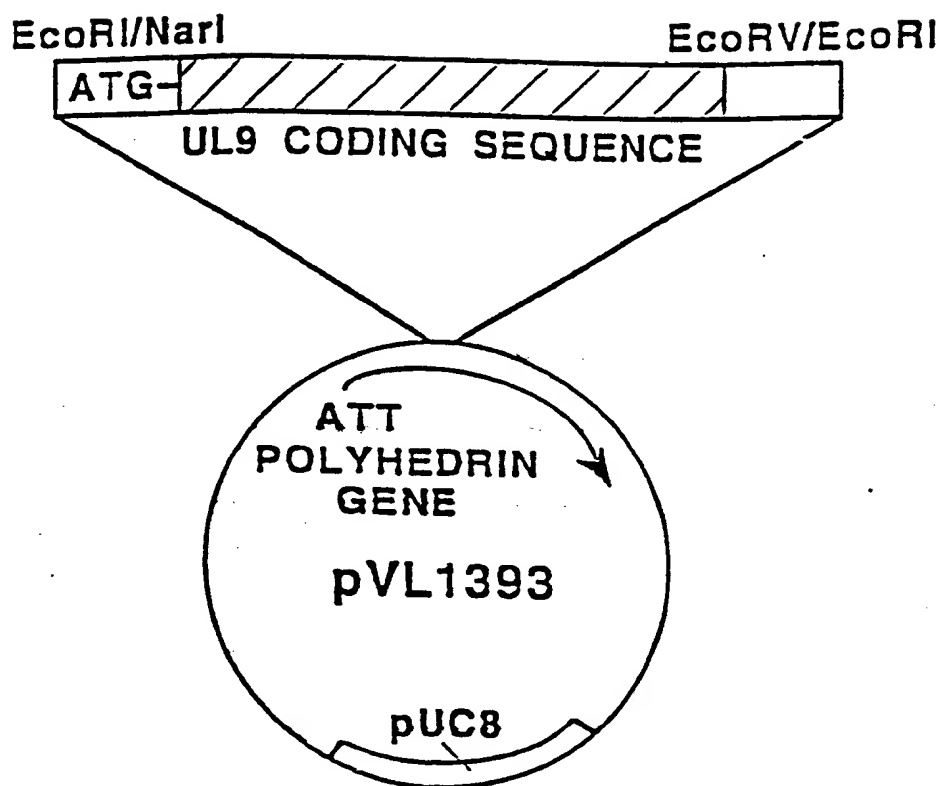


Fig. 7

7 / 39

Fig. 8

Fig. 8 GST-UL9 +
kD MARKER GST-UL9 +THROMBIN

kD MARKER GST-UL9

97.4 >

69.0 >

46.0 >

30.0 >

21.5 %

8 / 39

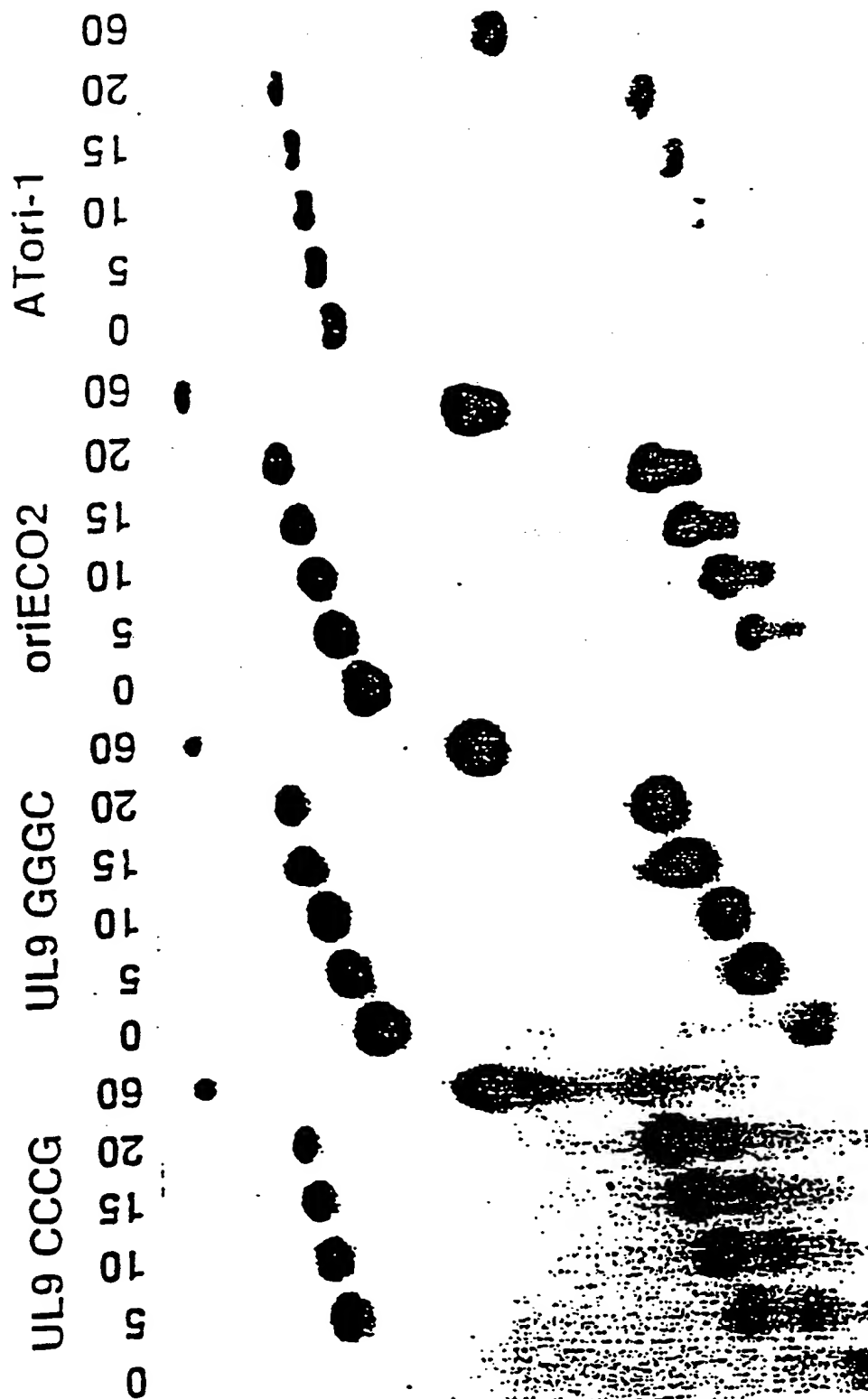


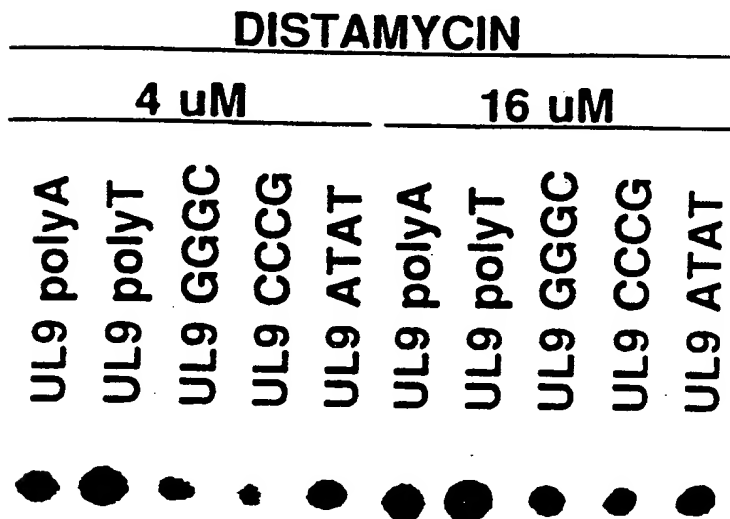
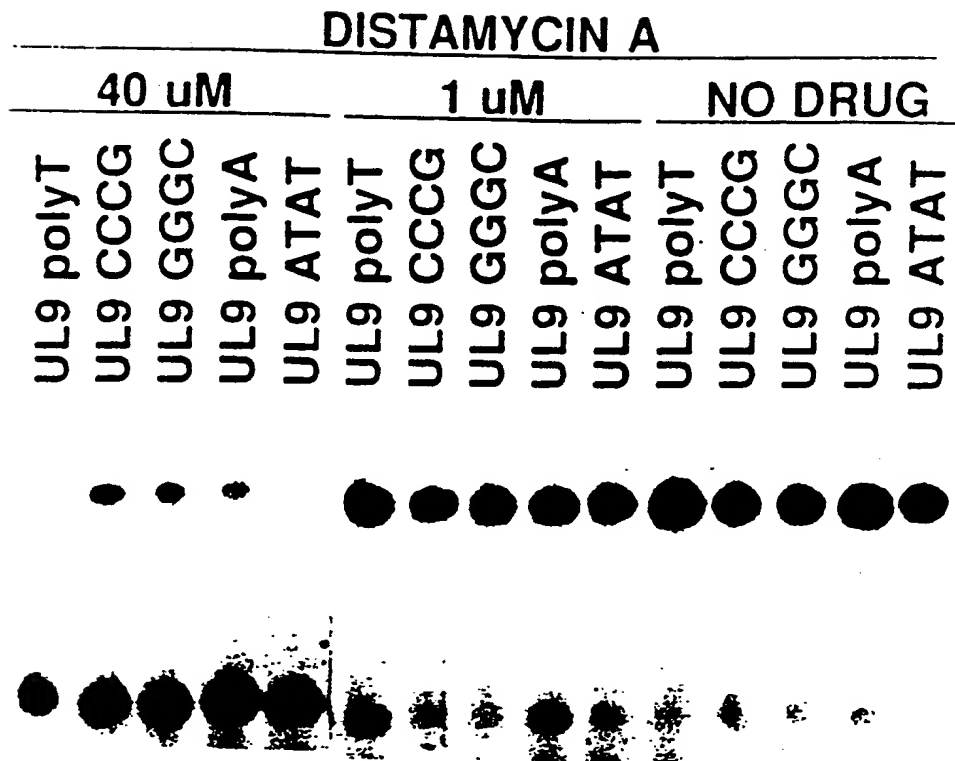
Fig. 9

9/39



Fig. 9 (con't)

10/39



11/39



●	UL9 polyA
●	UL9 polyT
●	UL9 ATAT
●	UL9 CCCG
●	UL9 GGGC
●	oriECO2
●	oriECO3
●	ATori-1
●	UL9 polyA
●	UL9 polyT
●	UL9 ATAT
●	UL9 CCCG
●	UL9 GGGC
●	oriECO2
●	oriECO3
●	ATori-1

NO DRUG

50 μ M ACTINOMYCIN D

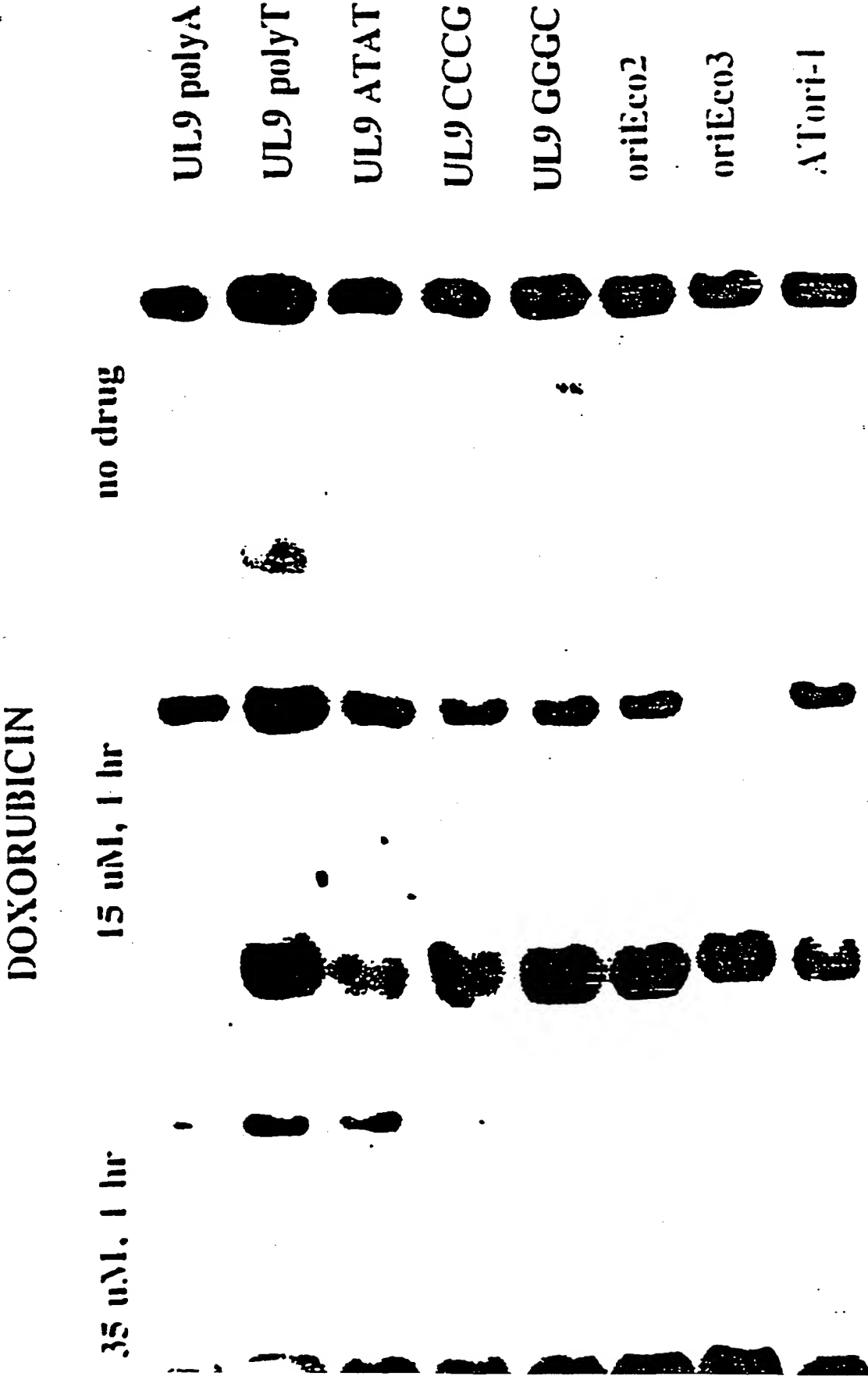


Fig. 10C

13/39

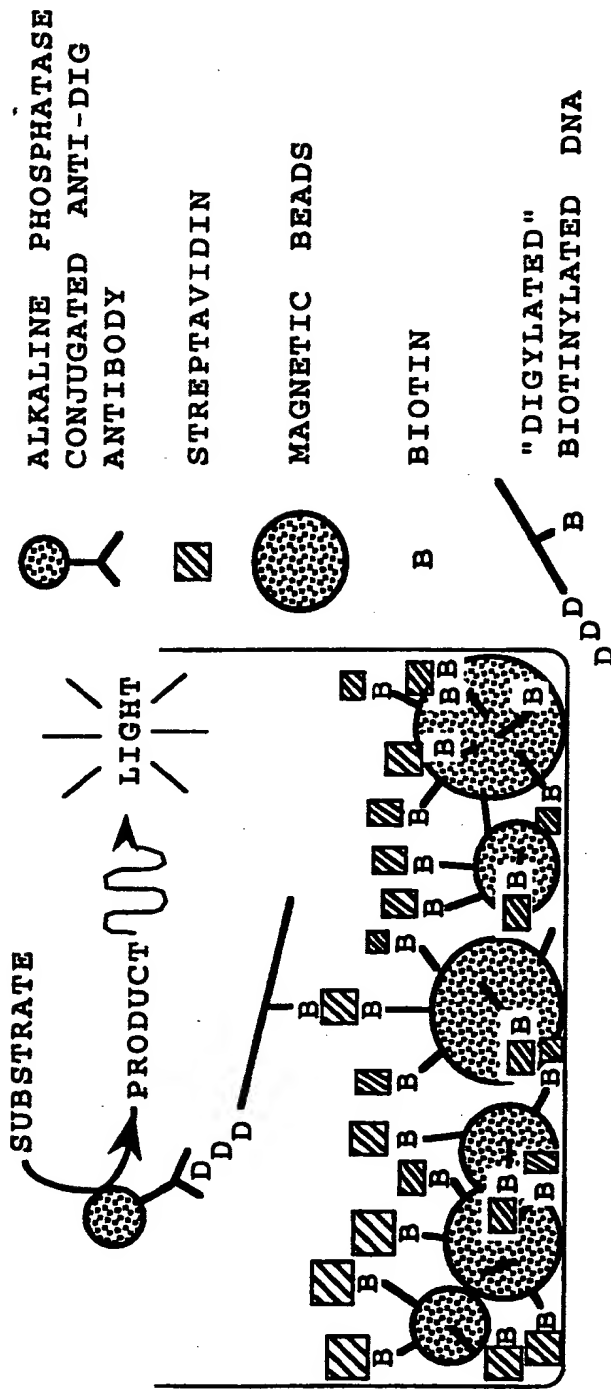


FIG. 11A

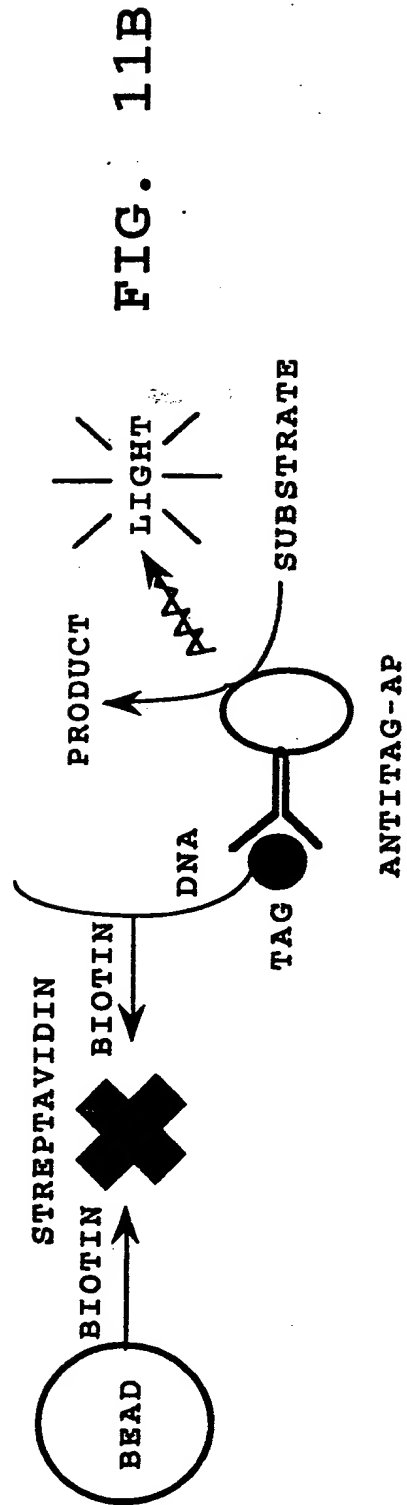


FIG. 11B

14/39

Table II. Demonstration Assay Test Matrix

Sequence	Test Mixtures							
	1	2	3	4	5	6	7	8
<u>AGCTTTCGCACCTTAGCT</u>	+	+	-	-	-	-	-	-
<u>AGCATTCGCACCTTAGCA</u>	+	-	-	+	+	-	-	-
<u>AGCCTTCGCACCTTAGCC</u>	+	-	-	-	-	-	-	+
<u>AGCGTTCGCACCTTAGCG</u>	+	-	-	-	-	-	-	-
<u>TGCTTTCGCACCTTTGCT</u>	+	-	-	-	-	-	+	-
<u>TGCAATTCGCACCTTTGCA</u>	+	-	-	+	+	-	+	-
<u>TGCCCTTCGCACCTTTGCC</u>	+	-	-	-	-	-	+	+
<u>TGCGTTCGCACCTTTGCG</u>	+	-	-	-	-	-	+	-
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
<u>CCATTCGCACCTTCCAT</u>	+	-	-	+	-	-	-	+
<u>CCCTTTCGCACCTTCCCT</u>	+	-	-	-	-	-	+	+
<u>CCGTTCGCACCTTCCGT</u>	+	-	-	-	-	-	-	+
<u>CCTTTTCGCACCTTCCTT</u>	+	-	-	-	-	-	+	+
:	:	:	:	:	:	:	:	:

Fig. 12

Fig 13

16 / 39

ACTT 032	CCAT 084	GACT 136	GTGT 188	TGTT 240
AGAA 033	CCCA 085	GAGA 137	GTTA 189	TTAA 241
AGAC 034	CCCC 086	GAGC 138	GTTT 190	TTAC 242
AGAG 035	CCCG 087	GAGG 139	GTTG 191	TTAG 243
AGAT 036	CCCT 088	GAGT 140	GTTT 192	TTAT 244
AGCA 037	CCGA 089	GATA 141	TAAA 193	TTCA 245
AGCC 038	CCGC 090	GATC 142	TAAC 194	TTCC 246
AGCG 039	CCGG 091	GATG 143	TAAG 195	TTCC 247
AGCT 040	CCGT 092	GATT 144	TAAT 196	TTCT 248
AGGA 041	CCTA 093	GCAA 145	TACA 197	TTGA 249
AGGC 042	CCTC 094	GCAC 146	TACC 198	TTGC 250
AGGG 043	CCTG 095	GCAG 147	TACG 199	TTGG 251
AGGT 044	CCTT 096	GCAT 148	TACT 200	TTGT 252
AGTA 045	CGAA 097	GCCA 149	TAGA 201	TTTA 253
AGTC 046	CGAC 098	GCCC 150	TAGC 202	TTTC 254
AGTG 047	CGAG 099	GCCG 151	TAGG 203	TTTG 255
AGTT 048	CGAT 100	GCCT 152	TAGT 204	TTTT 256
ATAA 049	CGCA 101	GCGC 154	TATA 205	
ATAC 050	CGCC 102	GCGA 153	TATC 206	
ATAG 051	CGCG 103	GCGG 155	TATG 207	
TATA 052	CGCT 104	GCGT 156	TATT 208	

Fig. 13 (con't)

17/39

GATC
GACT
GTCA
GTAC
GCTA
GCAT

AGTC
AGCT
ATCG
ATGC
ACTG
ACGT

TAGC
TACG
TGCA
TGAC
TCAG
TCGA

CGAT
CGTA
CATG
CAGT
CTAG
CTGA

Fig. 14A

Screening
Sequence

GCGTANXXXXCGTTCGCACTTXXXXCTTCGTCCCAAT
CGCATNXXXXGCAAGCGTGAAXXXXXGAAGCAGGGTTA

Test
Site

Test
Site

Fig. 14B

18/39

[Sort by average rank:]													
rank	longo	iseo	enc%	r918A	rank	%r918B	rank	%r1022A	rank	%r1022B	rank	ave r%	ave rank
1	246	TTCC		47	14	44	14	41	4	27	3	40	9
2	242	TTAC		49	19	47	19	55	17	48	15	49	18
3	188	TACC		57	43	48	22	50	7	39	10	48	21
4	206	TATC		50	20	48	18	54	16	59	45	52	25
5	7	7IAACG		56	38	56	52	48	6	34	4	49	25
6	247	ITTCG		56	36	49	28	58	20	54	34	54	29
7	254	ITTTG		51	21	35	4	66	46	63	59	54	33
8	27	27IACGG		55	30	55	49	63	33	51	24	56	34
9	202	TAGC		65	74	48	18	67	52	18	1	49	36
10	243	ITTAG		61	57	53	43	58	21	51	23	56	36
11	251	ITTGG		61	59	44	13	60	23	62	53	57	37
12	184	TAAC		66	83	51	35	45	5	52	30	54	38
13	3	3IAAAG		60	54	59	60	65	37	42	12	56	41
14	6	6IAACC		70	101	58	51	51	9	34	5	53	42
15	189	TACG		68	82	53	45	51	8	53	31	56	42
16	66	6CAAC		54	27	58	59	70	64	52	29	59	45
17	34	3IAGAC		55	32	52	41	63	32	67	75	59	45
18	2	2IAAAC		72	117	50	31	52	12	51	25	58	46
19	54	54IATCC		55	29	58	61	75	93	38	7	56	48
20	11	11AAGG		68	90	59	62	60	28	48	18	59	49
21	39	39IAGCG		49	18	53	44	80	138	39	8	55	52
22	38	38IAGCC		55	34	46	17	80	133	58	41	60	56
23	195	ITAAAG		70	105	63	72	57	19	54	32	61	57
24	248	ITTCT		70	104	52	40	65	38	60	46	61	57
25	28	28ACGC		58	45	49	25	78	116	59	44	61	58
26	22	22IACCC		64	72	49	28	65	40	71	93	62	58
27	58	58IATGC		63	67	65	77	68	53	55	37	63	59
28	43	43IAGGG		41	8	51	38	86	170	49	21	57	59
29	214	ITCCC		68	87	62	70	65	41	58	42	63	60
30	42	42IAGGC		43	8	49	24	90	185	44	14	56	60
31	207	ITATG		58	48	67	87	62	29	68	79	64	60
32	23	23IACCG		54	25	58	57	80	134	52	26	61	61
33	51	51IATAG		48	15	73	111	68	55	63	63	63	61
34	219	ITCGG		62	60	74	114	65	36	61	48	65	65
35	48	48IAGTC		18	2	42	9	74	80	80	160	54	65
36	249	TTGA		71	109	51	38	71	71	62	55	64	68
37	250	ITTCG		56	35	50	30	67	49	80	161	63	69
38	119	11CTCG		54	28	65	178	61	27	60	47	65	69
39	55	55IATCG		58	37	84	169	68	44	52	28	64	70
40	215	ITCCG		62	61	58	55	73	81	70	86	68	71
41	231	ITGCG		63	68	65	79	70	63	69	83	67	73
42	181	18GGAA		43	10	72	107	79	128	67	73	65	80
43	255	ITTTG		59	48	51	37	64	35	87	199	65	80
44	14	14IAATC		71	112	50	33	82	152	52	27	64	81
45	238	ITGTC		72	119	64	74	65	42	70	89	68	81
46	35	35IAGAG		58	41	55	47	87	184	62	54	65	82
47	241	ITTAA		78	137	84	170	51	10	41	11	63	82
48	18	18IACAC		68	80	50	29	97	216	36	8	62	83
49	47	47IAGTG		51	22	42	10	105	231	65	68	66	83
50	245	ITTCA		81	157	68	84	60	25	65	69	68	84
51	205	ITATA		70	102	68	85	71	70	69	82	69	85
52	210	ITCAC		65	75	45	15	21	2	57.3	250	43	86
53	38	38IAGAT		70	100	43	11	102	224	39	9	64	86
54	244	ITTTAT		80	149	62	69	61	28	71	99	69	86
55	234	ITGGC		56	40	61	68	82	150	70	88	67	87
56	256	ITTTT		54	28	58	58	71	67	86	195	67	87
57	183	183IAAA		74	124	74	117	68	54	63	61	70	89
58	203	203IAGG		84	178	59	63	62	30	70	90	69	90
59	63	63IATTG		60	51	78	122	80	137	63	56	70	92
60	227	227ITGAG		75	128	39	8	71	73	81	164	67	92
61	116	116CTAT		73	121	74	118	65	43	70	87	71	92
62	217	217ITCGA		60	53	85	175	69	59	71	95	72	96
63	59	59IATGG		62	65	82	150	79	130	56	39	70	96
64	62	62IATTC		64	69	71	102	93	203	43	13	68	97
65	185	185GTGA		44	12	62	71	76	105	87	201	67	97
66	61	61IATTA		77	140	87	179	67	50	49	22	70	98
67	50	50IATAC		54	24	71	97	105	232	56	38	71	98
68	235	235ITGGG		60	52	71	104	78	115	76	121	71	98
69	166	166GGCC		80	148	24	1	77	114	77	133	65	99

Fig. 15

19/39

70: 107:CGGG	65: 78:	103: 218:	67: 51:	81: 50:	73: 99
71: 232:TGCT	60: 97:	49: 27:	73: 79:	87: 198:	70: 100
72: 158:GCTC	33: 3:	66: 83:	71: 72:	131: 243:	75: 100
73: 19:ACAG	69: 95:	65: 75:	91:197:	54: 36:	70: 101
74: 10:AAAG	80: 151:	70: 83:	73: 82:	70: 85:	73: 103
75: 208:ITAT	70: 106:	72: 105:	75: 97:	73: 105:	72: 103
76: 1:AAAA	83: 171:	78: 137:	67: 47:	63: 58:	73: 103
77: 182:GTCC	78: 135:	60: 66:	70: 65:	78: 150:	71: 104
78: 157:GCTA	48: 18:	97: 209:	77:1111:	68: 81:	73: 104
79: 191:GTTG	110: 244:	43: 12:	78:102:	83: 60:	73: 105
80: 172:GGGT	38: 4:	58: 56:	89:192:	81: 168:	67: 105
81: 150:GCCC	58: 44:	80: 142:	88:172:	63: 62:	72: 105
82: 15:AAATG	68: 88:	70: 94:	108:237:	22: 2:	87: 105
83: 108:TAAT	94: 220:	67: 88:	66: 45:	67: 74:	74: 106
84: 187:GTGG	44: 11:	100: 220:	69: 58:	78: 140:	73: 107
85: 184:GTCT	82: 82:	48: 23:	77:1109:	100: 238:	72: 108
86: 115:CTAG	70: 98:	94: 203:	60: 24:	73: 108:	74: 108
87: 120:CTCT	81: 58:	114: 246:	59: 22:	73: 107:	77: 108
88: 167:GGCG	65: 73:	81: 145:	80:135:	68: 80:	73: 108
89: 239:GTGT	72: 114:	37: 5:	81:144:	82: 171:	68: 109
90: 233:TGGA	66: 84:	65: 78:	78:106:	82: 170:	72: 110
91: 52:ATAT	84: 70:	94: 199:	78:123:	81: 49:	74: 110
92: 220:TCGT	73: 122:	47: 21:	83:157:	78: 142:	70: 111
93: 183:GTCC	68: 81:	83: 158:	69: 57:	79: 146:	74: 111
94: 25:ACGA	60: 55:	74: 118:	100:221:	62: 52:	74: 111
95: 226:TGAC	71: 111:	73: 110:	71: 69:	80: 158:	74: 112
96: 33:AGAA	55: 33:	82: 155:	122:248:	48: 16:	77: 113
97: 114:CTAC	61: 58:	77: 125:	54: 15: 78.17	1253:	84: 113
98: 211:TCAG	75: 130:	70: 95:	75:1100:	78: 128:	74: 113
99: 104:CGCT	69: 93:	82: 184:	64: 34:	77: 131:	75: 113
100: 142:GATC	84: 68:	33: 3:	84:180:	92: 224:	68: 114
101: 230:TGCC	73: 123:	60: 67:	73: 77:	85: 188:	73: 114
102: 209:TCAA	72: 115:	69: 91:	84:158:	71: 82:	74: 114
103: 182:GGAC	41: 7:	83: 157:	67: 48:	149: 244:	65: 114
104: 87:CAAG	65: 184:	66: 82:	68:174:	49: 19:	72: 115
105: 252:TTGT	82: 162:	56: 50:	77:1113:	78: 138:	73: 115
106: 222:CTCT	89: 94:	77: 127:	84:183:	68: 77:	75: 115
107: 174:GGTC	65: 78:	107: 237:	54: 14:	77: 134:	78: 115
108: 30:ACTC	73: 120:	65: 81:	131:250:	48: 17:	79: 117
109: 71:ACAG	13: 1:	83: 182:	88:106:	76: 122:	65: 118
110: 45:AGTA	60: 50:	77: 128:	105:234:	64: 66:	77: 118
111: 201:TAGA	94: 216:	77: 129:	72: 75:	83: 57:	78: 118
112: 29:ACTA	81: 153:	42: 8:	131:249:	64: 67:	79: 118
113: 24:ACCT	58: 39:	83: 195:	95:210:	54: 35:	75: 120
114: 31:ACTG	68: 89:	73: 109:	112:240:	58: 43:	78: 120
115: 175:GGTG	71: 110:	94: 200:	55: 18:	80: 158:	75: 122
116: 131:GAAG	69: 91:	71: 100:	78:120:	83: 175:	75: 122
117: 229:TGCA	78: 138:	83: 183:	78:118:	67: 72:	78: 122
118: 163:GGAG	49: 17:	106: 234:	69: 60:	85: 184:	77: 124
119: 90:CCGC	83: 172:	82: 148:	69: 56:	75: 119:	77: 124
120: 253:TTTA	75: 128:	72: 108:	74: 85:	84: 180:	76: 125
121: 118:CTCC	78: 142:	57: 54:	75: 98:	89: 207:	75: 125
122: 151:GCCG	52: 23:	110: 243:	84:184:	67: 71:	78: 125
123: 155:GCGG	62: 84:	69: 92:	108:238:	75: 114:	79: 127
124: 221:TCCT	75: 134:	71: 98:	82:153:	77: 129:	78: 129
125: 179:GTAG	141: 250:	59: 84:	75: 95:	73: 108:	87: 129
126: 122:CTGC	89: 200:	76: 121:	74: 91:	74: 109:	78: 130
127: 197:ITACA	87: 192:	88: 183:	71: 68:	68: 78:	79: 130
128: 143:GATG	62: 63:	57: 53:	105:229:	84: 181:	77: 132
129: 170:GGGC	43: 9:	92: 180:	78:119:	89: 209:	75: 132
130: 44:AGGT	65: 77:	69: 90:	124:247:	75: 116:	83: 133
131: 57:ATGA	72: 118:	88: 181:	82:201:	57: 40:	77: 135
132: 13:AAATA	86: 186:	106: 235:	14: 1:	76: 120:	70: 136
133: 130:GAAC	90: 205:	65: 78:	80:136:	76: 125:	78: 136
134: 236:TTGT	75: 132:	60: 65:	78:122:	92: 225:	78: 136
135: 216:TCCT	83: 169:	71: 99:	82:149:	77: 130:	78: 137
136: 41:AGGA	81: 158:	51: 39:	151:253:	71: 97:	89: 137
137: 237:TGTA	82: 166:	74: 115:	74: 89:	83: 178:	78: 137
138: 9:AAAGA	87: 191:	78: 133:	82:154:	67: 70:	79: 137
139: 218:TCGC	72: 113:	78: 132:	75: 96:	90: 211:	79: 138
140: 56:ATCT	75: 133:	81: 146:	89:189:	69: 84:	79: 138

Fig. 15(con't)

20/39

141	108ICGGT	64	71	94	201	73	78	68	205	80	139
142	98ICGAC	94	219	84	188	54	13	80	159	78	140
143	145IGCAA	57	42	105	230	77	112	84	179	81	141
144	112ICGTT	80	152	82	151	78	101	81	165	80	142
145	96ICCTT	68	92	29	2	108	235	105	241	77	143
146	48IAGTT	70	107	47	20	118	243	87	200	80	143
147	171IGGGG	55	31	112	244	77	107	86	194	82	144
148	12IAAGT	78	138	89	184	102	223	54	33	80	145
149	154IGCGC	89	203	74	113	98	214	82	51	80	145
150	89ICCGA	95	222	50	32	74	88	104	239	81	145
151	240ITGTT	79	148	65	80	88	173	84	183	79	146
152	79ICATG	92	209	79	138	78	121	75	118	81	146
153	165IGGCA	80	49	105	232	72	74	95	231	83	147
154	204ITAGT	96	225	88	178	77	108	88	78	81	147
155	87ICCGG	85	182	82	152	73	80	83	174	81	147
156	200ITACT	97	230	84	187	74	88	74	110	82	148
157	132IGAAT	75	129	108	254	70	81	81	167	83	148
158	148IGCAC	59	47	115	248	87	183	75	115	84	148
159	49IATAA	85	79	84	163	128	248	72	102	87	148
160	4IAAAT	88	183	105	231	78	117	84	85	83	149
161	70ICACC	89	199	98	207	75	89	71	98	83	151
162	110ICGTC	98	233	92	191	73	83	71	98	83	151
163	28IACGT	83	168	55	48	148	252	78	138	90	152
164	8IAACT	81	160	83	159	94	182	77	128	81	152
165	82ICCGT	81	159	77	130	73	84	102	238	84	153
166	5IAACA	83	170	83	197	84	181	71	91	83	155
167	84ICCAT	108	242	54	48	87	177	80	155	82	155
168	190IGTTC	84	174	88	98	85	188	85	192	80	156
169	180IGCTT	44	13	89	188	92	202	92	222	79	156
170	78ICATC	82	213	83	161	79	128	78	124	83	156
171	123ICTGG	88	185	89	185	81	145	74	111	83	157
172	228ITGAT	102	240	71	101	81	148	78	143	83	158
173	135IGACG	82	184	79	135	89	191	79	147	82	158
174	149IGCCA	70	99	97	210	105	228	72	101	88	160
175	128ICTTT	79	144	77	128	85	189	88	203	82	161
176	159IGGTG	40	5	104	229	88	171	105	240	84	161
177	17IACAA	82	187	112	245	98	213	49	20	85	161
178	225ITGAA	91	208	78	124	81	140	83	178	83	162
179	184IGGAT	67	86	100	221	82	148	85	191	84	162
180	178IGTAC	92	212	78	120	85	188	79	149	83	162
181	77ICATA	155	252	80	34	102	222	78	139	96	162
182	105ICGGA	84	173	110	242	52	11	92	223	84	162
183	53IATCA	75	131	83	158	105	230	77	135	85	163
184	40IAGCT	75	127	71	103	104	228	87	197	84	163
185	173IGGTA	71	108	98	208	88	187	80	154	83	164
186	158IGCGT	74	125	78	123	119	245	81	162	86	164
187	88ICCGT	68	193	80	138	74	92	101	237	86	165
188	32IACTT	89	98	82	154	124	251	81	188	92	167
189	144IGATT	67	85	80	143	100	220	91	220	85	167
190	223ITCTG	77	139	85	204	77	110	91	217	85	168
191	134IGACC	97	229	89	86	87	180	82	173	84	168
192	75ICAGG	98	229	73	108	57	182	79	153	84	168
193	91ICCGG	89	150	81	147	81	141	97	233	85	168
194	72ICACT	93	215	107	238	33	156	83	84	87	168
195	153IGCGA	72	118	100	219	111	239	72	103	89	169
196	192IGTTT	125	248	74	119	84	159	79	151	91	169
197	137IGAGA	89	201	40	7	103	225	248.59	248	78	170
198	127ICTTG	81	158	92	193	82	151	85	185	85	171
199	83ICCTA	88	187	92	192	78	103	88	204	88	172
200	178IGGTT	117	246	70	88	81	142	88	202	89	172
201	213ITCCA	77	141	78	134	87	178	99	234	85	172
202	80IATGT	79	145	83	160	116	242	78	141	89	172
203	125ICTTA	88	189	52	42	95	211	79.5	254	78	174
204	168IGGCT	101	239	85	172	78	129	80	157	88	174
205	139IGAGG	92	210	94	198	85	187	78	123	87	175
206	147IGCAG	70	103	120	251	90	198	79	152	90	176
207	37IAGCA	79	143	81	144	94	207	89	210	86	176
208	181ICGCA	79	147	88	182	78	124	65.76	252	82	176
209	224ITCTT	93	214	73	112	84	185	91	216	85	177
210	183ICGGG	91	207	116	250	82	31	91	219	90	177
211	180IGTAT	179	254	100	218	75	94	79	144	108	177

Fig. 15(con't)

21/39

212:	64:ATTT	84: 175:	83: 165:	94:208:	82: 172:	88:	180:
213:	141:GATA	95: 224:	82: 153:	91:200:	79: 145:	87:	181:
214:	140:GAGT	117: 245:	63: 73:	96:215:	85: 189:	90:	181:
215:	169:GGGA	82: 165:	145: 254:	89:180:	75: 113:	98:	181:
216:	97:CGAA	153: 251:	101: 223:	24: 3:	798: 246:	269:	181:
217:	94:CCTC	82: 211:	82: 149:	86:175:	85: 190:	86:	181:
218:	186:GTGC	205: 255:	356: 256:	74: 87:	77: 127:	178:	181:
219:	82:CCAC	97: 232:	85: 174:	80:131:	86: 183:	87:	183:
220:	113:CTAA	85: 180:	109: 240:	70: 62: *60.31	251:	88:	183:
221:	212:TCAT	89: 188:	77: 131:	89:183:	91: 215:	87:	184:
222:	65:CAAA	84: 179:	116: 249:	95:209:	72: 100:	92:	184:
223:	99:CGAG	103: 241:	99: 215:	65: 39:	180: 245:	112:	185:
224:	102:CGCC	88: 185:	106: 233:	70: 68: *35.62	249:	88:	186:
225:	78:ICAGT	96: 227:	85: 177:	93:205:	78: 137:	88:	187:
226:	126:CTTC	81: 181:	90: 187:	81:147: *88.2	256:	84:	188:
227:	21:ACCA	85: 181:	99: 213:	113:241:	75: 117:	93:	188:
228:	189:GTTA	178: 253:	210: 255:	81:145:	72: 104:	135:	189:
229:	69:ICACA	82: 183:	83: 184:	98:219:	90: 212:	88:	190:
230:	111:CGTG	81: 155:	108: 239:	81:136:	93: 227:	91:	190:
231:	138:GAGC	84: 178:	80: 139:	91:189: *171.29	247:	85:	190:
232:	100:CGAT	89: 202:	80: 141:	163:255:	81: 163:	103:	190:
233:	74:ICAGC	95: 223:	94: 202:	78:125:	90: 213:	89:	191:
234:	108:CGGC	99: 238:	98: 211:	93:204:	75: 112:	91:	191:
235:	129:GAAA	87: 180:	93: 196:	87:185:	87: 196:	89:	192:
236:	181:GTCA	98: 234:	98: 212:	76:104:	91: 218:	91:	192:
237:	73:GAGA	97: 231:	102: 228:	98:217:	71: 94:	92:	192:
238:	20:ACAT	89: 186:	91: 188:	109:238:	79: 148:	92:	193:
239:	83:CCAG	90: 204:	85: 171:	96:212:	85: 188:	89:	193:
240:	152:GCCT	88: 194:	91: 189:	94:208:	85: 187:	89:	194:
241:	95:CCTG	89: 197:	85: 173:	87:178:	99: 235:	90:	196:
242:	148:GCAT	94: 218:	100: 217:	98:218:	77: 132:	92:	196:
243:	81:CCAA	91: 208:	101: 224:	88:188:	83: 177:	91:	199:
244:	177:GTAA	130: 249:	133: 252:	79:127:	81: 169:	106:	199:
245:	117:CTCA	81: 154:	101: 222:	89:194:	95: 230:	91:	200:
246:	108:CGTA	120: 247:	115: 247:	72: 78:	97: 232:	101:	201:
247:	121:CTGA	101: 238:	95: 205:	87:181:	84: 182:	92:	202:
248:	18:AAAT	96: 226:	80: 140:	117:244:	89: 206:	95:	204:
249:	88:CCCC	84: 177:	103: 228:	91:198:	93: 226:	93:	207:
250:	85:CCCA	99: 237:	102: 225:	83:155:	115: 242:	100:	215:
251:	188:GTGT	217: 256:	136: 253:	80:132:	92: 221:	131:	216:
252:	124:CTGT	94: 221:	96: 208:	106:233:	89: 208:	98:	218:
253:	133:GACA	108: 243:	110: 241:	87:177:	90: 214:	99:	219:
254:	138:GACT	88: 188:	103: 227:	159:254:	93: 228:	110:	224:
255:	68:CAAT	99: 235:	99: 214:	105:227:	94: 229:	99:	226:
256:	80:CATT	94: 217:	87: 180:	348:256: *838.83	255:	178:	227:

Fig. 15 (con't)

22/39

Distamycin: Rank vs average rank based on 4 experiments and ideal rank

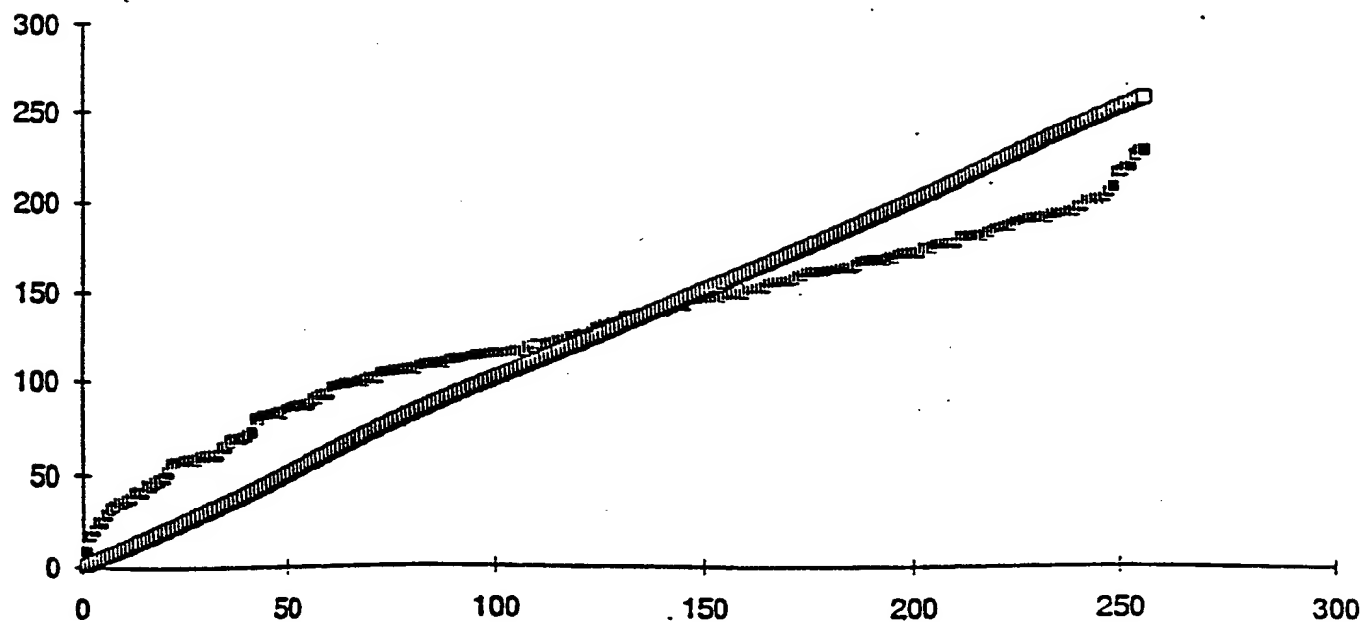


Fig. 16

Distamycin: rank vs r% score

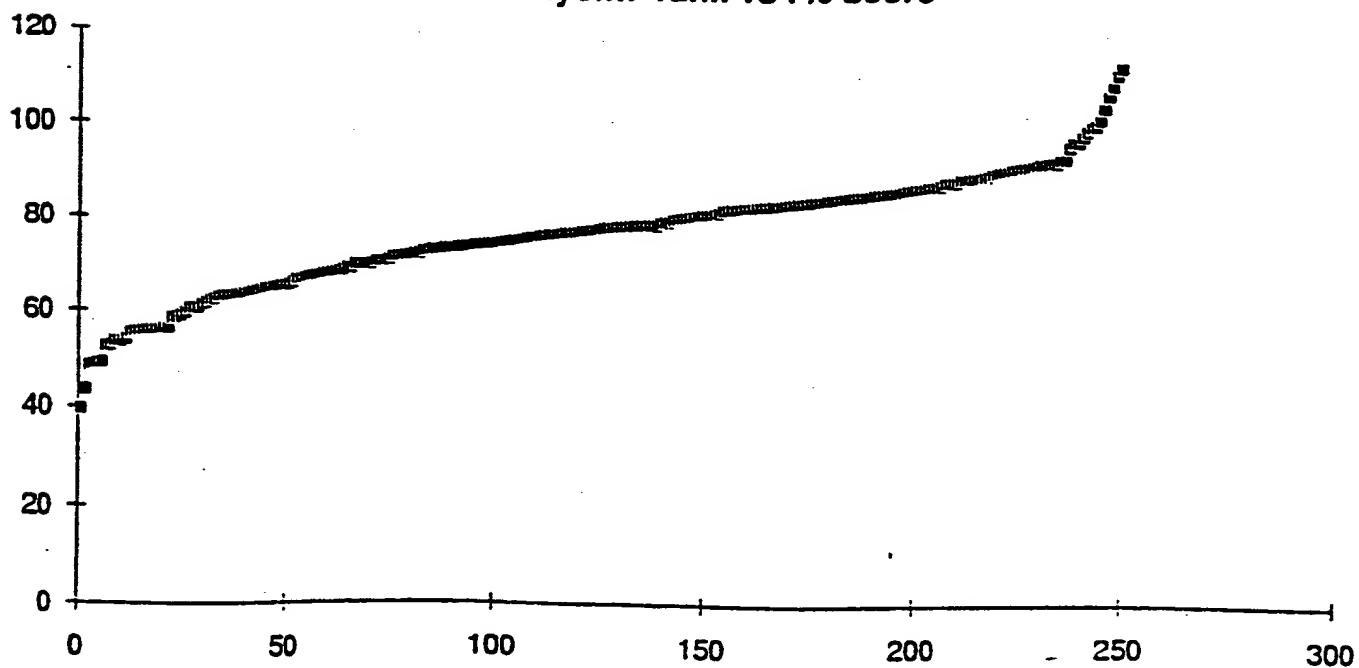


Fig. 17

Aclomycin D Screens

Aclomycin D screens		Screen A		Screen B		Screen C		Screen D		Screen E		Screen F		Screen G		Screen H		Average % Bound		Average Rank		Rank Order		olgo	
Rank	Order	%	rank	%	rank	%	rank	%	rank	%	rank	%	rank	%	rank	%	rank	%	%	Rank	Order				
1	155	60.57	59	46.11	4	42.29	23	44.02	4	38.20	8	34.00	2	29.96	4	41.60	13.25	1	155	1					
2	153	39.38	5	58.55	13	55.55	42	53.50	23	45.91	17	34.13	3	23.44	2	44.83	14.25	2	153	2					
3	145	50.93	23	62.15	18	40.63	45	54.21	45	53.49	37	34.10	4	23.44	3	47.90	18.88	3	145	3					
4	158	60.46	57	67.86	32	52.76	22	43.71	20	27.70	1	38.75	5	30.84	6	45.95	20.43	4	158	4					
5	181	40.61	6	58.30	9	56.74	38	52.16	46	51.43	31	38.12	6	29.18	3	48.40	20.50	5	181	5					
6	176	44.18	13	68.99	30	58.07	24	42.54	43	45.76	15	52.39	8	38.98	12	50.98	23.38	6	176	6					
7	146	59.55	40	67.11	31	53.97	39	52.23	18	49.68	25	31.57	30	31.08	16	50.44	23.50	7	146	7					
8	148	48.01	20	44.74	3	47.41	37	37.09	10	72.50	113	46.77	1	40.33	5	46.98	23.50	8	148	8					
9	149	41.26	8	71.02	57	57.71	52	56.10	13	45.84	16	45.61	14	38.90	11	51.15	24.88	9	149	9					
10	123	41.41	9	47.23	5	52.90	15	38.33	27	45.98	18	63.29	95	54.08	51	50.03	29.50	10	123	10					
11	150	62.68	67	71.19	58	47.09	36	51.80	15	39.32	9	48.13	21	44.99	25	52.31	29.63	11	150	11					
12	154	52.56	28	75.16	59	68.03	11	35.58	12	42.24	12	45.75	15	26.09	10	50.74	30.50	12	154	12					
13	157	64.05	79	67.88	47	54.73	20	31.55	32	47.56	20	48.91	26	47.67	30	52.62	30.63	13	157	13					
14	170	59.34	47	69.04	47	64.46	51	48.86	28	49.88	26	44.85	12	44.06	22	54.33	31.50	14	170	14					
15	170	38.78	4	74.60	81	55.26	22	50.57	31	67.01	61	36.91	6	44.71	23	52.55	32.57	15	170	15					
16	147	60.36	54	76.98	126	51.43	13	48.35	27	41.47	3	N/A	N/A	9	N/A	42.29	32.57	16	147	16					
17	160	65.45	86	60.43	17	44.44	21	43.40	36	60.80	65	49.31	9	43.52	20	48.68	34.63	17	160	17					
18	152	66.55	91	68.49	12	55.18	25	46.77	38	59.72	61	39.45	7	39.31	15	53.34	34.63	18	152	18					
19	181	57.38	40	71.27	60	55.18	47	53.13	24	49.55	24	55.07	58	51.67	39	56.44	35.57	19	181	19					
20	157	59.75	51	75.10	87	61.83	39	53.52	25	53.54	38	43.75	11	39.02	13	55.41	36.63	20	157	20					
21	184	45.31	16	74.70	83	50.42	10	58.34	45	N/A	N/A	48.20	22	N/A	N/A	56.60	38.67	21	184	21					
22	184	43.98	12	84.23	168	50.42	7	53.34	34	52.36	34	54.94	37	44.90	24	56.60	39.63	22	184	22					
23	168	57.76	44	77.87	114	60.88	49	55.48	28	53.87	40	48.64	16	46.27	28	56.92	43.50	23	168	23					
24	182	55.81	35	64.55	23	61.23	36	50.92	58	62.93	71	59.85	70	54.50	52	59.58	47.38	24	182	24					
25	194	44.69	15	68.08	37	37.70	3	60.31	30	73.92	123	58.25	45	56.29	61	56.88	47.63	25	194	25					
26	164	56.47	36	95.76	233	66.07	23	47.44	56	57.80	3	35.46	4	43.03	17	51.95	47.88	26	164	26					
27	164	60.37	55	77.48	112	51.38	26	55.13	34	62.79	34	52.36	39	49.89	33	58.29	48.00	27	164	27					
28	174	72.42	134	59.19	15	62.16	40	53.34	60	63.83	35	55.54	39	50.30	35	59.04	49.50	28	174	28					
29	183	59.48	48	72.78	72	65.90	54	60.50	56	62.28	41	54.49	34	50.30	35	60.82	54.00	29	183	29					
30	180	51.75	26	51.20	7	67.33	41	53.34	55	57.10	55	59.60	48	42.89	46	57.05	54.50	30	180	30					
31	178	60.01	52	63.68	22	53.91	68	37.20	14	79.06	13	47.47	19	58.88	48	56.39	55.13	31	178	31					
32	178	71.11	126	70.88	55	71.18	89	33.44	8	67.92	51	48.78	25	45.74	27	58.16	55.75	32	178	32					
33	182	41.58	10	31.15	1	66.84	58	33.44	65	56.34	88	61.63	80	50.58	36	57.22	59.75	33	182	33					
34	182	59.24	46	75.14	88	61.63	38	49.93	76	69.51	42	63.97	105	55.24	57	61.10	60.25	34	182	34					
35	182	61.32	51	39.91	2	160.43	20	49.60	6	57.04	54	58.85	62	52.10	40	65.43	60.50	35	182	35					
36	185	60.56	58	68.82	46	67.37	63	N/A	92	67.95	82	57.63	52	48.79	32	63.21	60.71	36	185	36					
37	172	68.78	106	51.13	6	N/A	N/A	N/A	158	62.38	68	45.43	13	39.11	14	57.54	60.83	37	172	37					
38	172	68.78	106	51.13	6	N/A	N/A	N/A	158	62.38	68	45.43	13	39.11	14	57.54	60.83	38	172	38					
39	173	53.97	31	75.77	93	62.54	43	34.55	8	45.87	8	N/A	43	57.53	69	58.02	61.29	39	173	39					
40	173	70.86	120	72.53	69	54.30	19	45.96	51	55.29	44	N/A	N/A	N/A	120	61.28	71.13	40	173	40					
41	190	65.87	89	62.58	20	56.80	26	50.71	97	36.16	7	72.85	166	51.58	38	60.34	73.00	41	190	41					
42	190	68.19	102	70.11	51	65.99	55	38.55	11	49.41	111	63.68	98	58.47	75	62.43	72.00	42	190	42					
43	184	77.84	176	78.45	118	69.99	74	53.25	66	67.75	203	53.17	31	54.69	53	63.21	72.13	43	184	43					
44	183	62.35	65	66.98	29	62.47	42	61.83	220	48.32	107	41.67	8	63.60	121	62.42	75.13	44	183	44					

Fig. 18(1)

Aclhomycin D Screens

15	88	CGAC	43.00	11	76.73	102	N/A	N/A	66.67	122	57.17	28	77.94	155	59.62	68	53.99	48	62.16	76.29	45	98
16	91	CGCG	37.67	3	77.48	111	N/A	N/A	60.54	69	52.81	14	N/A	N/A	79.40	222	52.95	44	60.14	77.17	46	91
17	243	TTAG	51.43	25	62.56	19	105.85	217	54.93	46	68.07	68	75.78	134	60.31	71	53.08	45	66.49	78.13	47	243
18	168	GGCC	63.70	76	83.98	164	55.83	23	73.29	176	53.29	16	47.93	22	60.94	74	61.57	101	62.57	81.50	48	168
19	244	TTAT	67.43	97	54.61	8	78.26	128	60.11	64	73.07	109	79.43	165	58.05	58	48.32	31	64.66	82.50	49	244
20	198	TAT	78.47	179	66.80	28	51.75	14	N/A	N/A	69.98	79	82.49	185	56.19	44	55.14	55	65.83	83.43	50	198
21	82	CGAC	73.84	144	83.47	159	86.64	187	55.34	51	43.07	5	61.87	67	54.65	36	43.50	19	62.85	83.50	51	82
22	338	CGAC	65.53	68	84.11	165	71.20	91	60.34	66	58.70	33	41.87	11	69.61	152	57.12	67	63.56	84.15	52	338
23	71	CAOB	67.02	94	69.66	49	57.96	29	62.29	81	71.29	88	56.54	52	57.97	55	83.33	228	65.76	94.50	53	71
24	202	TAGC	81.74	206	83.44	158	74.12	110	30.93	4	57.63	29	51.92	33	53.76	32	62.78	111	62.04	85.38	54	202
25	250	TTGC	75.37	156	87.05	186	72.08	97	29.08	3	71.76	96	84.95	73	57.02	49	46.82	29	63.01	86.13	55	250
26	97	ACGG	66.61	92	80.03	133	71.62	93	67.25	126	45.40	7	61.44	180	57.66	53	44.58	7	63.08	86.38	56	97
27	12	ATGA	77.49	168	90.13	210	62.87	45	63.06	86	73.29	111	51.51	32	47.38	18	43.87	21	63.70	86.38	57	12
28	132	GAAT	73.50	141	86.93	184	62.25	41	56.35	53	77.91	153	33.80	3	59.09	64	53.79	59	63.20	87.25	58	132
29	251	TTGG	69.22	111	58.03	14	N/A	N/A	77.04	77	75.50	128	79.60	166	62.10	84	52.61	42	65.71	88.66	59	251
30	186	GTCG	50.43	22	77.20	105	N/A	N/A	77.04	77	75.50	128	79.60	166	62.10	84	52.61	42	65.71	88.66	60	186
31	100	GTAT	69.01	108	77.53	113	N/A	N/A	62.19	79	70.37	81	51.35	30	61.06	75	66.70	143	65.46	89.86	61	100
32	30	ACTC	63.85	78	71.43	61	76.78	133	73.84	180	72.55	104	55.77	47	61.40	78	53.99	49	66.20	91.25	62	30
33	4	AAAT	73.87	146	68.75	44	63.52	46	N/A	N/A	68.50	71	80.07	168	59.68	69	63.55	119	68.33	94.71	63	4
34	140	GAGT	65.50	87	68.60	43	81.50	168	68.10	131	55.38	22	72.01	106	69.75	154	56.56	62	67.17	96.63	64	140
35	31	ACTG	77.20	167	73.83	77	17.29	221	54.07	44	71.45	93	69.14	58	56.10	42	59.14	79	70.95	97.63	65	31
36	118	GTAC	61.36	62	71.74	63	93.90	205	64.94	103	60.33	41	69.14	89	66.03	126	61.09	98	68.97	98.38	66	118
37	175	GGTG	68.31	104	86.21	181	110.54	220	68.74	137	63.73	45	55.58	48	48.69	24	45.23	26	68.42	98.50	67	175
38	7	AACG	68.40	105	83.42	156	72.35	99	69.29	142	72.53	103	53.85	39	61.67	81	56.74	64	67.28	98.63	68	7
39	28	CATG	57.47	41	95.73	231	N/A	N/A	62.70	82	59.31	35	72.84	N/A	56.67	46	58.61	164	68.75	99.63	69	28
40	235	TTTG	71.02	122	102.27	247	64.12	50	59.40	92	59.82	39	72.84	114	63.90	102	57.10	66	68.77	100.13	70	235
41	39	ACTA	85.30	85	71.48	62	85.39	185	63.66	92	82.22	179	51.02	29	61.85	93	60.07	87	67.62	100.25	71	39
42	288	TTTT	73.85	145	88.88	200	72.92	104	60.98	71	60.55	42	76.19	138	56.85	47	55.48	58	68.21	100.63	72	288
43	37	CATA	45.90	17	76.72	101	63.84	47	62.16	78	67.89	67	74.86	N/A	74.86	200	76.16	199	66.79	101.29	73	37
44	83	CCAG	31.15	1	58.14	11	N/A	N/A	63.10	87	67.89	176	67.36	81	71.18	174	71.90	179	63.46	101.29	74	83
45	247	TTGG	69.06	109	60.09	16	166.33	231	68.62	136	74.92	122	59.22	60	57.85	54	60.23	90	77.04	102.25	75	247
46	75	CACT	52.35	27	68.33	39	69.51	73	84.51	223	62.64	47	73.72	122	70.33	165	64.84	129	68.28	103.15	76	75
47	93	CGTA	51.13	24	89.50	207	70.81	85	65.38	107	N/A	N/A	47.85	21	74.08	194	59.98	85	65.53	103.29	77	93
48	18	ACAC	91.71	246	72.96	70	N/A	N/A	61.60	74	63.79	49	77.56	152	60.48	73	57.57	70	69.32	104.86	78	18
49	86	CGTC	67.50	98	87.74	192	80.15	160	35.55	10	75.52	129	N/A	N/A	55.86	41	N/A	N/A	67.05	105.00	79	86
50	198	TACG	79.63	189	93.31	154	67.73	65	58.29	54	53.62	17	75.92	135	64.41	110	63.37	117	68.28	105.13	80	198
51	168	GTGT	66.82	93	65.88	27	N/A	N/A	60.60	70	83.11	186	76.88	144	67.53	137	59.33	80	68.56	105.29	81	168
52	131	GAAG	71.58	129	56.81	10	77.86	146	162.35	244	69.45	75	67.31	80	64.55	113	54.04	50	77.99	105.88	82	131
53	58	ATGC	79.53	187	82.93	151	103.10	216	63.78	94	61.71	44	46.36	19	51.19	29	62.86	113	68.93	106.63	83	58
54	10	AAGC	85.40	222	89.00	201	80.81	163	61.87	75	N/A	N/A	57.84	57	47.53	20	36.08	9	65.48	106.71	84	10
55	102	CGCC	63.63	75	78.60	120	71.74	94	69.04	140	85.17	199	63.39	72	57.14	50	N/A	N/A	69.82	107.14	85	102
56	179	GTAG	69.21	110	80.15	134	70.28	77	88.42	233	72.97	106	62.48	69	63.71	99	50.09	34	69.66	107.75	86	179
57	128	CTTC	70.58	116	75.24	90	73.93	108	61.26	72	72.59	105	53.06	36	77.14	214	63.83	122	68.44	107.88	87	128
58	253	TTTA	78.85	182	79.98	132	71.25	90	63.10	88	67.09	63	73.29	118	54.50	35	67.89	156	69.49	108.00	88	253
59	19	ACAG	77.59	173	74.85	86	78.05	149	62.96	85	70.40	92	77.05	146	58.01	56	60.51	92	69.93	108.63	89	19
60	56	CAAG	66.52	90	75.24	91	90.63	196	60.15	65	68.83	70	68.56	87	73.52	189	59.47	82	70.36	108.75	90	56
61	25	ACGA	47.26	19	71.84	64	77.06	137	62.81	84	84.75	196	N/A	N/A	58.05	57	77.52	209	68.47	109.43	91	25

Fig. 18(2)

92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138		
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138		
109.93	110.13	111.00	111.50	112.00	113.50	113.57	114.75	114.75	116.00	116.13	116.75	116.75	117.00	117.83	117.83	117.75	118.17	118.38	119.38	119.38	120.43	120.57	121.25	121.88	122.50	122.50	122.71	124.00	124.38	124.37	125.13	125.29	125.98	127.13	127.68	128.25	129.25	129.38	130.00	130.13	130.25	130.75	131.25	131.29	131.63	131.63		
69.96	71.23	69.93	74.32	70.61	69.86	69.17	69.10	71.33	70.28	70.93	72.97	68.74	67.28	70.63	70.22	68.61	71.25	71.03	70.07	70.63	74.87	71.74	70.93	71.88	73.94	72.21	72.71	72.34	69.49	71.71	74.36	72.21	72.28	71.10	73.00	70.11	71.23	71.96	72.19	76.62	76.93	72.50	71.87	72.05	75.34			
71	139	203	194	43	94	99	88	187	76	200	130	169	188	108	223	104	204	145	116	173	96	97	134	131	95	109	83	47	177	150	116	65	166	73	231	221	102	216	54	103	128	74	100	144	157	249		
57.60	66.08	76.62	75.29	52.93	60.78	61.10	57.22	74.43	58.50	76.19	64.94	70.04	74.45	62.28	80.63	61.75	76.66	66.99	71.02	61.07	59.99	63.61	65.74	65.74	60.87	62.33	59.66	53.17	71.74	67.45	63.20	57.09	69.02	84.81	80.17	61.68	78.89	55.14	61.75	64.70	61.18	66.85	67.96	116.99				
40	59	76	129	119	91	62.89	33	243	185	94	202	111	130	181	65	86	139	96	156	107	61	63	136	51	125	153	108	182	135	225	118	93	160	142	178	218	109	248	90	23	115	233	188	164				
55.55	58.37	61.56	66.22	65.38	62.89	70.92	53.90	66.84	72.41	63.23	75.22	64.44	72.05	72.95	59.23	82.57	67.59	63.63	69.59	64.18	58.75	58.34	57.31	57.21	65.94	69.97	64.29	70.31	67.43	80.50	65.98	70.13	67.14	68.34	71.98	78.36	64.32	107.68	62.89	46	103	64.07	84.34	73.41	70.28			
168	150	78	205	163	163	64	64	68.12	59.82	76.59	108.76	48.53	32.30	2	92	N/A	186	63	28	N/A	215	86	137	177	5	14	109	10	95	104	153	93	115	43	97	27	75	101	59.12	55.61	71.50	73.07	109	209	218			
80.31	77.45	N/A	68.86	88.58	79.33	N/A	60.45	68.12	59.82	76.59	108.76	48.53	32.30	2	92	N/A	186	63	28	N/A	215	86	137	177	5	14	109	10	95	104	153	93	115	43	97	27	75	101	59.12	55.61	71.50	73.07	109	209	218			
108	90	55	85	73	204	31	152	69	218	62	133	95	185	77	149	211	72	84	224	130	N/A	N/A	N/A	165	229	N/A	N/A	98	157	121	236	209	125	110	182	144	230	99	234	170	145	101	177	123	37			
73.07	70.03	65.72	70.82	69.31	68.06	58.16	77.89	68.39	90.38	70.84	71.51	83.02	89.62	77.51	77.51	87.65	69.14	70.74	93.49	75.69	66.61	N/A	76.75	80.07	94.26	71.87	78.55	74.72	110.06	87.81	87.60	73.09	82.51	77.00	95.96	95.96	71.94	103.91	80.63	77.01	81.86	72.23	74.96	59.49				
212	95	165	123	144	35	118	17	80	57	157	60	106	177	93	59	112	98	105	130	N/A	205	224	62	241	228	226	232	90	N/A	128	111	184	6	63	63	32	155	120	215	213	161	19	164	167	1			
79.84	63.80	71.95	66.73	69.57	51.58	66.31	39.69	62.26	59.73	71.36	59.20	63.23	71.63	63.70	59.19	65.94	63.95	65.19	67.89	N/A	N/A	76.02	64.69	84.56	59.53	105.72	84.48	84.99	88.12	63.29	N/A	67.64	74.68	33.20	59.98	59.70	71.34	69.59	79.89	71.56	42.19	71.81	72.19	27.69				
80	210	59	226	114	119	204	178	57	76	132	81	37	60	196	N/A	87	117	157	64	203	165	92	34	78	27	165	189	180	N/A	88	175	171	219	158	83	141	61	170	170	138	N/A	123	69					
70.55	96.25	68.91	128.91	75.32	74.68	93.34	83.24	66.77	70.19	76.77	70.93	61.35	66.38	86.01	N/A	71.00	74.94	76.82	83.28	70.35	57.51	80.99	71.31	70.39	57.51	80.99	71.31	80.96	83.67	71.00	82.81	82.21	108.71	70.74	70.69	70.69	77.35	77.35	77.35	77.35	77.35	77.35	77.35	68.49				
127	36	95	25	53	24	97	237	45	80	79	66	135	141	109	235	76	116	117	153	108	254	144	172	107	138	149	75	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	214				
79.03	68.07	75.88	64.80	70.43	64.60	76.19	96.71	68.77	74.43	74.26	72.04	85.93	70.48	77.40	95.99	73.69	78.08	78.39	83.20	70.23	138.30	81.97	82.81	82.68	73.49	68.36	71.87	76.05	98.27	98.27	84.27	82.26	85.87	82.26	85.87	82.26	85.87	82.26	85.87	82.26	85.87	82.26	85.87	82.26	85.87			
70	112	121	32	140	183	103	142	154	174	63	39	214	138	18	14	71	238	50	83	169	166	33	64	64	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	101			
82.90	69.82	70.88	54.90	73.16	78.98	68.24	73.64	75.03	77.60	61.96	57.14	82.90	72.94	46.56	44.48	63.10	59.62	59.71	64.95	77.50	70.46	70.70	70.70	64.43	79.34	55.42	61.96	56.59	62.71	56.92	74.68	49.60	57.49	77.57	60.85	83.46	80.64	79.80	73.71	72.36	70.76	70.20	71.10	67.88				
99	108	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80	187	245	62	165	203	199	50	246	130	31	114	32	137
99	104	133	221	201	107	139	169	82	129	224	87	147	84	241	100	121	84	17	189	73	70	318	124	125	101	59	117	118	167	191	222	28	249	177	80													

Actinomycin D Screens

139	110	CGTC	75.37	158	90.14	211	60.78	32	64.62	99	78.01	135	72.24	110	70.41	167	66.33	142	71.99	131.75	139	110
140	205	206 TATA	76.38	165	78.80	125	64.00	48	68.41	133	78.22	155	81.07	175	N/A	N/A	65.13	132	73.14	133.29	140	205
141	219	219 TGGG	75.75	162	78.79	124	77.75	144	85.57	228	80.79	172	61.80	66	64.05	106	58.22	72	72.84	134.25	141	219
142	54	54 ATCC	87.49	233	85.52	175	N/A	144	55.49	50	84.75	197	69.39	90	62.58	88	62.26	107	72.50	134.29	142	54
143	3	3 AAGC	74.56	151	138.01	255	71.76	95	72.23	168	69.81	78	76.46	141	65.79	124	56.63	63	78.28	134.38	143	3
144	252	252 TGTG	74.55	150	90.30	215	75.85	122	65.62	109	N/A	N/A	75.12	129	59.40	66	67.54	152	72.63	134.71	144	252
145	186	186 AGTC	72.75	136	81.56	143	80.28	161	65.45	108	71.20	87	87.74	204	65.73	123	64.30	125	73.63	135.88	145	186
146	208	208 TATT	74.35	147	81.52	142	80.55	111	73.26	175	73.70	115	81.39	179	69.39	151	67.87	155	72.09	136.50	146	208
147	185	185 GTGA	88.36	236	73.98	78	70.11	75	65.10	104	73.70	115	81.39	179	69.39	151	67.87	155	72.09	136.50	147	208
148	148	148 GTGA	59.16	45	94.17	228	83.61	179	51.86	37	114.85	238	56.26	50	68.56	147	70.16	170	74.63	136.75	148	185
149	149	149 CCCTG	60.41	56	72.75	71	78.13	154	66.06	115	64.97	52	89.82	199	86.12	239	84.49	127	73.66	137.29	149	149
150	150	150 CCCTG	33.81	2	83.99	163	N/A	154	77.77	204	77.10	148	78.77	158	67.56	138	73.88	186	73.48	138.25	151	242
151	151	151 TTAC	71.05	123	68.46	42	76.73	131	74.29	182	77.10	148	78.77	158	67.56	138	73.88	186	73.48	138.25	151	242
152	152	152 ATAG	90.67	243	75.72	92	72.47	100	78.06	206	78.40	155	N/A	N/A	64.56	114	56.07	60	73.70	138.71	152	39
153	153	153 ATAG	77.83	175	67.82	35	74.69	115	72.74	172	87.62	210	80.45	170	61.13	76	68.56	163	73.87	139.50	153	11
154	154	154 AAGC	79.67	190	67.30	189	70.90	88	71.17	152	74.11	118	83.70	187	59.03	92	55.19	56	75.24	140.00	154	9
155	155	155 AAGA	67.08	95	93.18	223	72.20	98	64.04	97	74.11	118	83.70	187	59.03	92	55.19	56	75.24	140.00	154	9
156	156	156 ATTC	82.96	215	72.33	68	72.83	103	74.96	188	88.16	214	N/A	N/A	58.40	60	69.09	138	73.67	140.50	155	53
157	157	157 CGGA	63.61	74	98.11	241	39.09	228	63.49	91	125.14	240	56.74	53	62.74	89	62.54	110	83.93	140.75	156	103
158	158	158 TATT	79.23	185	90.15	212	59.59	30	74.94	185	74.35	119	74.77	128	68.54	146	64.20	124	73.22	140.88	158	207
159	159	159 TATG	93.62	249	87.02	195	85.34	52	98.45	134	75.90	134	102.67	220	55.02	117	50.50	37	76.08	141.00	159	51
160	160	160 GACG	80.79	198	78.50	119	82.57	174	61.50	73	76.79	142	86.71	201	69.99	157	61.99	105	74.17	141.00	160	135
161	161	161 TCAA	75.37	198	61.32	140	60.93	33	78.22	193	70.49	83	75.33	131	59.32	150	76.10	201	73.85	141.13	161	309
162	162	162 ATAT	60.38	241	86.41	196	68.38	72	70.27	145	71.98	100	74.35	125	66.20	128	64.05	123	73.08	141.25	162	20
163	163	163 TGCC	63.40	73	73.36	74	84.01	182	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	67.55	236	77.08	141.25	163	230
164	164	164 ATCC	69.83	113	84.20	167	360.75	232	40.14	16	67.00	60	81.62	183	70.05	158	76.84	205	106.30	142.00	164	22
165	165	165 CGCT	82.57	212	74.62	82	94.87	207	72.49	171	78.21	154	73.70	121	65.50	85	63.06	114	75.25	143.25	165	104
166	166	166 TACA	80.55	195	68.11	195	64.01	49	84.73	225	65.51	53	76.00	136	65.44	121	71.63	175	74.50	143.63	166	197
167	167	167 TACA	80.18	192	93.97	225	N/A	N/A	71.52	160	72.97	107	70.64	98	70.07	159	70.96	81	73.62	143.67	167	5
168	168	168 GTCC	53.05	29	79.84	131	65.24	208	69.04	139	90.32	217	70.64	98	70.07	159	70.96	172	73.62	144.13	168	183
169	169	169 CAAA	62.52	66	78.65	123	82.95	176	71.47	159	73.47	112	77.24	147	71.90	177	75.43	196	74.20	144.50	169	55
170	170	170 ATCA	77.99	177	69.12	203	73.53	106	62.74	83	76.35	135	72.15	108	65.43	120	84.40	230	75.21	145.38	170	53
171	171	171 CAAG	72.17	131	107.82	249	76.00	125	68.08	114	71.31	90	N/A	N/A	69.88	195	67.72	154	75.85	145.43	171	57
172	172	172 TATG	77.57	171	89.46	206	62.84	44	68.08	189	60.27	40	N/A	139	76.75	210	69.56	167	73.59	145.75	172	206
173	173	173 ATAT	80.96	201	75.78	94	53.13	17	70.42	147	76.36	137	83.92	188	74.06	195	74.52	189	73.64	146.00	173	16
174	174	174 CGCC	125.02	255	72.32	67	80.36	162	67.56	127	97.53	231	N/A	N/A	63.89	101	59.86	84	80.93	146.71	174	103
175	175	175 GTTT	72.98	139	87.83	193	72.04	96	58.86	58	89.18	216	72.37	112	83.53	231	65.86	133	75.31	147.25	175	128
176	176	176 TACT	87.43	232	65.63	65	67.96	65	79.20	209	80.78	171	80.56	172	74.16	196	61.91	106	74.70	147.25	176	200
177	177	177 TCGT	87.10	230	83.88	128	83.88	181	68.97	125	72.39	102	79.15	162	64.48	112	N/A	N/A	76.15	148.57	177	220
178	178	178 TCAAG	81.49	204	90.60	217	79.92	158	69.02	113	75.78	132	74.83	129	67.31	134	62.85	112	74.85	149.75	178	211
179	179	179 TCTG	72.42	135	95.29	230	91.12	198	74.14	181	69.37	74	75.15	130	63.92	103	67.58	153	76.13	150.50	179	223
180	180	180 AATT	84.54	220	69.46	48	78.00	148	66.50	121	77.27	147	74.02	124	75.56	204	77.01	207	75.31	152.38	180	16
181	181	181 CGTG	75.03	195	82.01	145	68.12	68	68.50	135	76.35	162	73.27	117	73.68	191	99.19	246	77.39	152.38	181	111
182	182	182 GTTA	70.81	119	81.09	139	73.96	109	67.71	129	71.71	N/A	N/A	N/A	71.71	176	99.86	247	77.52	153.17	182	125
183	183	183 ATGT	80.44	194	84.45	171	44.43	4	70.75	150	82.79	183	75.41	132	50.44	247	67.00	146	74.46	153.38	183	60
184	184	184 TCAAC	63.73	77	76.53	98	79.92	159	36.31	12	104.85	235	92.68	212	78.24	217	78.90	217	76.36	153.38	184	210
185	185	185 GATA	95.02	250	74.84	85	91.21	199	N/A	N/A	74.53	120	73.29	119	76.49	208	60.61	93	78.00	153.43	185	141

Fig. 18(4)

Actinomycin D Screens

186	34 AGAC	78.50	180	101.87	215	71.83	163	71.47	94	79.05	161	67.18	132	68.46	150	77.14	154.25	186	34
187	204 TAGT	91.18	245	92.03	201	72.26	169	66.99	59	89.40	91	68.99	149	68.61	165	78.70	156.00	187	204
188	181 GTCA	114.04	254	100.90	214	61.90	76	84.12	190	74.78	127	62.57	87	67.42	149	81.14	156.50	188	181
189	148 GATC	72.41	133	87.57	192	70.76	151	75.33	126	88.29	197	70.13	161	60.39	91	76.49	156.83	189	142
190	116 CTAT	80.41	193	77.22	139	N/A	N/A	70.87	86	76.94	145	71.01	172	64.33	126	76.71	156.71	190	116
191	42 AGCG	72.91	137	N/A	N/A	79.09	208	87.53	205	87.36	56	85.93	238	63.20	115	75.29	157.00	191	42
192	1 AAAA	87.21	231	77.71	143	68.37	132	87.37	205	N/A	N/A	66.09	127	66.30	141	75.95	157.29	192	1
193	314 GAGC	77.53	170	148.58	229	69.47	143	83.93	189	77.31	149	70.25	163	72.45	182	83.41	157.38	193	314
194	204 TTTC	80.58	196	N/A	N/A	66.43	119	76.70	140	80.53	171	63.79	100	70.39	171	75.48	158.00	194	204
195	117 CTCA	86.48	228	76.84	134	64.64	100	98.97	232	81.15	176	68.49	145	61.01	96	77.67	159.13	195	117
196	138 GACT	80.80	199	75.66	120	72.33	170	80.16	166	70.97	100	73.65	190	58.85	77	79.75	159.38	196	138
197	115 CTAG	75.57	161	77.38	140	69.02	139	N/A	N/A	N/A	N/A	67.27	133	66.00	137	76.40	159.50	197	115
198	72 CACT	75.54	159	N/A	N/A	75.76	188	75.12	124	N/A	N/A	61.70	82	85.97	232	76.50	159.83	198	72
199	15 AATG	90.31	240	70.79	84	71.42	158	76.96	143	76.95	160	68.18	141	83.78	135	76.74	160.00	199	15
200	143 GATG	82.18	209	82.35	172	75.18	187	N/A	N/A	58.13	49	70.94	170	83.55	229	75.34	160.00	200	143
201	91 ATTG	88.48	227	75.97	124	71.18	153	73.55	113	N/A	N/A	60.48	72	91.80	241	78.17	160.14	201	91
202	35 AGAG	80.89	200	75.97	124	71.18	153	73.55	113	N/A	N/A	60.48	72	91.80	241	78.17	160.14	202	35
203	85 CCGA	88.24	235	81.43	167	64.52	98	80.52	188	81.53	182	75.82	206	78.70	214	77.23	162.50	203	85
204	38 CAAT	70.31	115	80.88	164	59.51	55	73.65	114	N/A	N/A	72.00	180	78.82	215	76.93	163.43	204	38
205	14 AATC	90.09	239	68.88	71	70.66	148	77.69	150	81.51	181	71.16	173	58.53	161	76.47	164.25	205	14
206	348 TTCT	67.84	100	90.96	197	N/A	148	81.08	174	78.27	157	72.36	184	66.14	140	77.11	164.75	206	348
207	182 GTTC	76.24	164	132.12	227	64.83	102	84.50	192	84.60	190	109.00	249	74.60	190	82.80	165.14	207	182
208	113 AATA	93.48	248	84.97	164	68.82	124	84.98	198	86.32	199	75.68	205	75.46	198	83.20	167.75	208	113
209	36 CCCC	65.16	84	93.13	222	77.01	197	80.61	199	64.99	74	73.22	187	76.55	202	81.02	169.63	209	36
210	55 ATCG	88.83	237	121.85	224	70.70	149	80.06	164	N/A	N/A	65.49	122	67.16	148	82.38	170.29	210	55
211	109 CGTA	74.47	148	63.92	53	76.48	194	92.66	221	N/A	N/A	68.45	143	71.90	178	98.08	170.43	211	109
212	327 TGAG	62.76	69	73.79	107	77.12	200	88.01	203	85.01	193	73.72	192	71.94	181	78.04	171.13	212	327
213	127 CTTC	67.24	96	77.04	136	76.83	185	80.43	167	N/A	N/A	77.03	213	68.53	162	77.52	171.57	213	127
214	16 ATCT	79.60	168	67.66	212	95.73	238	75.40	127	76.35	140	67.91	140	60.20	89	81.46	172.13	214	16
215	8 AACT	63.39	72	78.72	152	133.37	243	N/A	N/A	82.01	164	71.98	179	71.48	174	84.36	173.14	215	8
216	35 AGCG	55.62	34	90.52	195	N/A	N/A	94.55	227	82.91	201	78.60	219	73.52	185	81.78	173.57	216	35
217	217 TCGA	74.55	149	70.83	82	80.91	218	87.39	208	76.72	143	74.91	201	78.18	210	78.87	174.88	217	217
218	113 CTAA	81.74	207	74.43	113	73.65	179	79.46	163	116.48	225	68.47	144	81.07	224	82.21	175.25	218	113
219	237 TGTA	76.22	178	120.76	223	80.70	217	71.33	91	79.83	167	80.22	224	80.00	220	83.23	175.50	219	237
220	215 TGAA	61.19	202	84.02	163	85.37	227	77.75	151	69.80	94	76.65	209	72.75	183	78.87	175.75	220	215
221	85 CACA	78.67	181	74.72	116	69.10	141	82.23	180	85.55	194	75.46	203	N/A	N/A	79.59	176.14	221	85
222	112 CGTT	76.02	163	78.20	160	77.13	201	85.76	201	69.94	96	77.00	212	N/A	N/A	79.44	177.57	222	112
223	88 CCGT	64.56	82	72.52	102	89.96	234	82.17	178	92.41	211	132.23	251	76.37	206	86.80	178.13	223	88
224	23 AGCG	92.45	247	83.22	177	72.11	168	79.16	160	75.57	137	72.05	182	86.09	233	80.06	178.50	224	23
225	2 AAAA	86.91	229	77.79	145	78.12	192	85.50	200	81.04	174	70.95	171	78.86	136	78.86	178.75	225	2
226	212 CAT	82.25	210	74.42	112	81.54	219	79.31	161	85.85	195	64.89	116	71.99	176	80.08	179.13	226	212
227	5 AACA	82.03	208	76.87	135	71.35	156	71.29	89	95.87	214	N/A	N/A	77.15	208	82.37	179.43	227	5
228	47 AGTG	68.89	107	89.02	151	78.89	207	82.81	184	81.22	177	83.88	232	75.15	192	80.36	180.63	228	47
229	214 TCCC	88.19	234	77.05	134	76.96	166	75.69	131	N/A	N/A	89.31	246	78.35	212	80.52	181.86	229	214
230	215 TCGG	90.57	242	82.42	173	73.62	178	76.58	138	89.79	207	80.86	227	68.31	159	80.30	182.38	230	215
231	38 AGCG	86.33	226	77.43	110	80.43	216	82.60	189	78.21	156	77.19	215	69.96	168	80.06	183.50	231	38
232	234 TCGC	64.41	80	N/A	N/A	72.86	173	82.32	181	84.98	192	86.57	241	80.18	222	79.97	183.86	232	234

Fig. 18(5)

Achromycin D Screens

233	15 AGTA	71.40	128	77.89	115	81.38	168	77.24	202	88.91	215	104.60	222	84.62	234	78.19	211	83.03	186.63	233	45
234	238 TGA1	85.84	223	87.88	194	76.05	129	79.93	214	N/A	N/A	76.86	159	70.36	168	91.45	239	81.25	186.71	234	238
235	213 TCCA	91.13	244	101.29	245	61.15	35	78.36	210	N/A	N/A	85.98	198	74.48	199	75.26	193	81.24	188.86	235	213
236	108 CGGT	83.16	217	90.19	213	73.16	105	104.88	240	93.95	225	79.38	184	70.58	188	75.10	191	83.60	190.38	236	108
237	229 TCCA	71.23	127	87.23	187	79.15	155	87.74	231	87.87	213	77.48	151	86.70	242	79.41	218	82.10	190.50	237	229
238	17 AGCA	83.10	216	85.57	176	87.39	190	68.12	116	90.83	220	N/A	N/A	74.37	197	79.91	219	81.04	190.57	238	17
239	209 TGTG	76.41	166	94.15	227	N/A	N/A	77.07	199	77.32	148	110.58	224	84.72	235	67.09	147	83.91	192.29	239	209
240	231 TGGG	74.69	153	86.37	182	87.30	189	91.97	237	80.94	173	80.57	173	78.96	221	78.64	213	82.43	192.50	240	231
241	225 TGAC	84.64	221	89.18	204	106.84	218	76.03	191	94.57	228	73.50	120	76.82	211	67.50	151	83.63	193.00	241	225
242	236 TGGT	79.11	184	76.66	100	97.31	211	N/A	N/A	81.30	175	102.37	219	82.29	229	87.08	235	85.59	193.29	242	236
243	232 TGGT	57.69	43	95.24	229	91.54	200	82.10	221	84.29	191	91.86	210	83.35	230	83.14	227	83.65	193.88	243	232
244	14 AGCT	96.57	251	84.93	173	76.14	127	73.03	174	90.46	219	84.28	189	74.46	198	87.02	234	83.36	195.63	244	14
245	4 AGCA	81.23	203	93.99	226	77.53	142	91.86	236	93.02	223	68.22	84	113.69	250	90.63	238	88.75	200.25	245	4
246	1 AGGT	83.59	219	80.67	136	94.43	206	65.83	110	95.29	229	104.27	221	86.47	240	92.25	242	87.95	200.38	246	1
247	43 AGGG	86.24	225	97.24	239	N/A	N/A	66.15	117	87.41	207	77.87	154	79.81	223	91.72	240	83.78	200.71	247	43
248	2 ACCA	135.86	255	97.64	240	74.90	118	76.00	190	84.70	195	86.68	200	80.99	228	71.92	180	88.59	200.88	248	2
249	233 TEGA	71.07	124	110.33	252	92.16	202	79.48	211	N/A	N/A	77.27	148	87.17	244	82.74	226	85.75	201.00	249	233
250	40 AGCT	87.89	99	87.27	188	95.79	209	86.76	229	103.52	233	93.53	213	89.04	245	75.44	197	87.26	201.63	250	40
251	12 AGCT	81.66	205	92.86	221	74.22	111	107.95	242	N/A	N/A	89.87	208	71.18	175	118.83	250	90.94	201.71	251	12
252	46 AGTT	82.33	211	96.83	238	87.79	193	74.41	183	83.59	187	91.39	178	78.87	220	81.71	225	83.36	204.38	252	46
253	239 TGGG	75.55	160	88.41	197	98.72	213	96.03	239	85.88	202	84.86	191	85.53	237	88.22	237	87.90	209.50	253	239
254	238 TGTG	85.93	224	88.81	199	78.81	153	77.37	203	118.62	239	N/A	N/A	77.66	216	103.38	248	90.08	211.71	254	238
255	216 TCGT	100.46	253	95.84	234	72.52	101	91.36	235	92.72	222	88.88	206	76.37	207	94.84	243	89.13	212.63	255	216
256	240 TGTG	82.60	213	110.20	251	77.91	147	86.95	230	84.68	194	87.10	202	85.39	238	95.58	245	88.80	214.75	256	240

Fig. 18(6)

29 / 39

Actinomycin D Screens

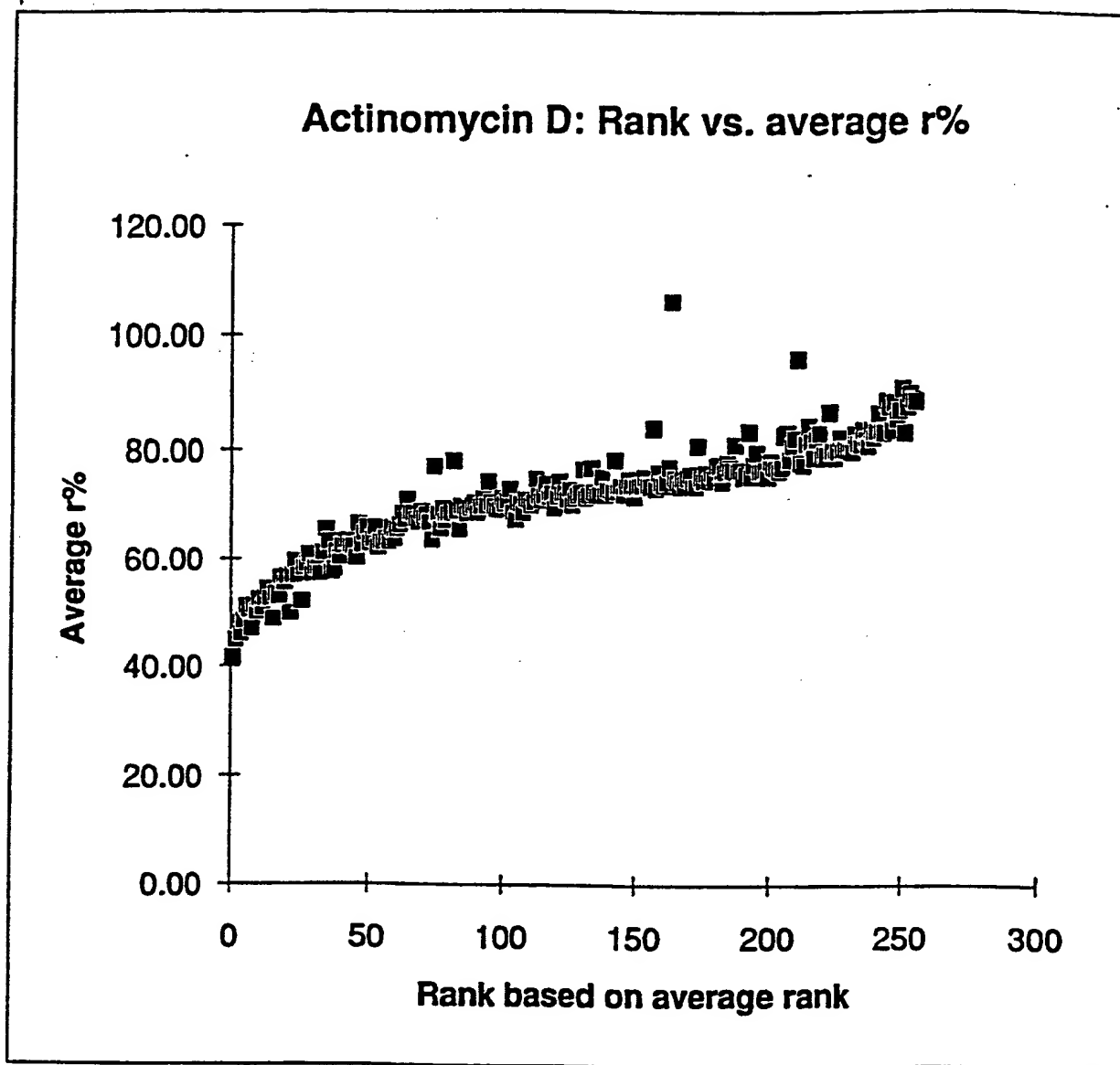
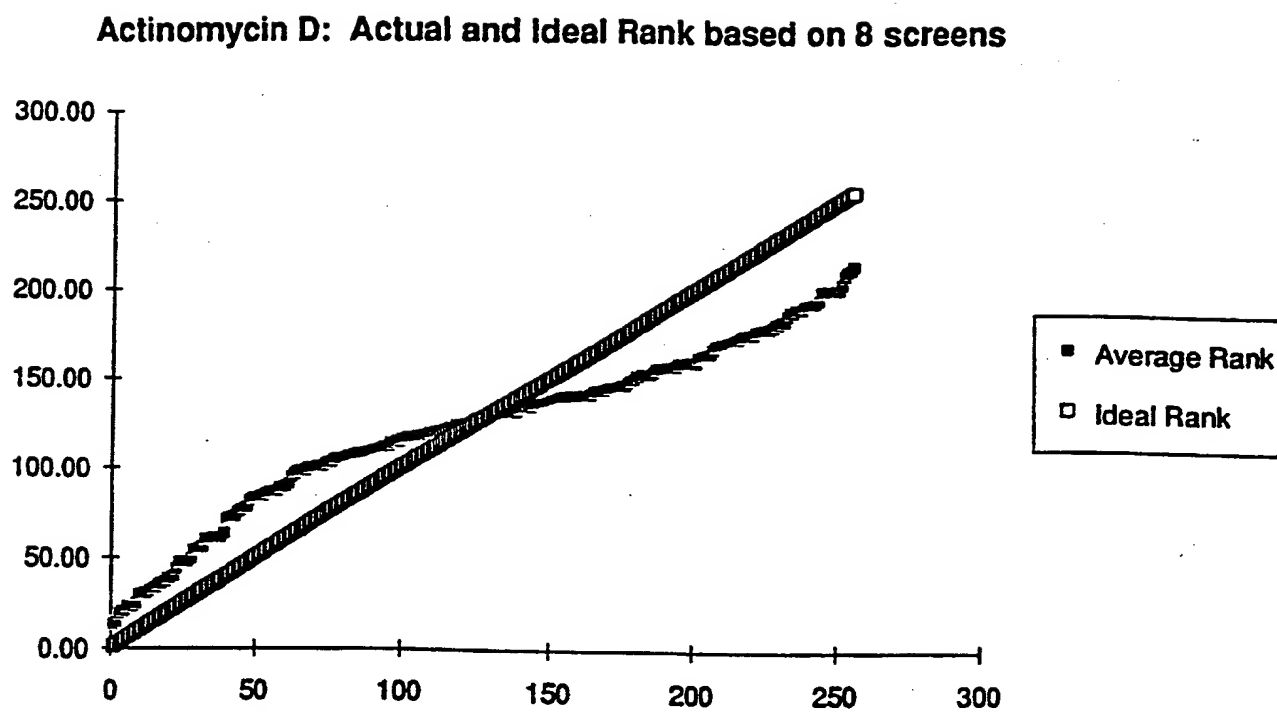


Fig. 19

30 / 39

**Fig. 20**

31 / 39

Actinomycin D:	
Variance from mean	
A1	-4.38
C1	1.09
G1	7.71
T1	-4.42
A2	-0.29
C2	4.39
G2	-4.02
T2	-0.08
A3	2.29
C3	-4.88
G3	2.86
T3	-0.27
A4	-0.92
C4	2.01
G4	0.56
T4	-1.65

Fig. 21

32 / 39

	N1N2	N1N3	N1N4	N2N3	N2N4	N3N4	XNX	XX
AA	-3.39	-0.94	-4.21	0.80	-2.00	2.56	-1.47	-0.01
AC	-3.00	-10.57	-2.04	-3.61	2.28	1.72	-4.14	-1.63
AG	-10.85	-2.35	-3.13	2.14	0.89	1.29	-0.73	-2.47
AT	-0.25	-3.64	-8.12	-0.49	-2.33	3.58	-2.98	0.95
CA	1.52	1.61	-1.63	7.80	3.78	-3.33	2.69	2.00
CC	2.14	-2.89	4.97	-4.43	4.63	-5.45	0.87	-2.58
CG	-1.38	5.44	2.58	8.94	5.72	-3.56	5.58	1.33
CT	2.08	0.20	-1.55	5.26	3.44	-7.19	1.82	0.05
GA	-1.48	9.81	6.63	1.08	-3.47	-1.31	3.17	-0.57
GC	23.80	2.40	7.64	-7.41	-1.51	9.80	0.44	8.73
GG	10.42	11.66	7.28	-4.76	-4.87	5.21	3.39	3.62
GT	-1.89	6.96	9.30	-5.00	-6.22	-2.27	0.37	-3.05
TA	2.19	-1.33	-4.47	-0.54	-1.99	-1.60	-1.66	0.02
TC	-5.36	-8.47	-2.54	-4.07	2.63	1.95	-2.92	-2.49
TG	-14.27	-3.31	-4.46	5.12	0.52	-0.69	-1.39	-3.28
TT	-0.26	-4.59	-6.22	-0.84	-1.49	-0.72	-3.04	-0.61
stdev:	8.44	6.13	5.54	4.86	3.52	4.22	2.78	3.04

Fig. 22

33 / 39

Actinomycin D: Cumulative dinucleotide analysis

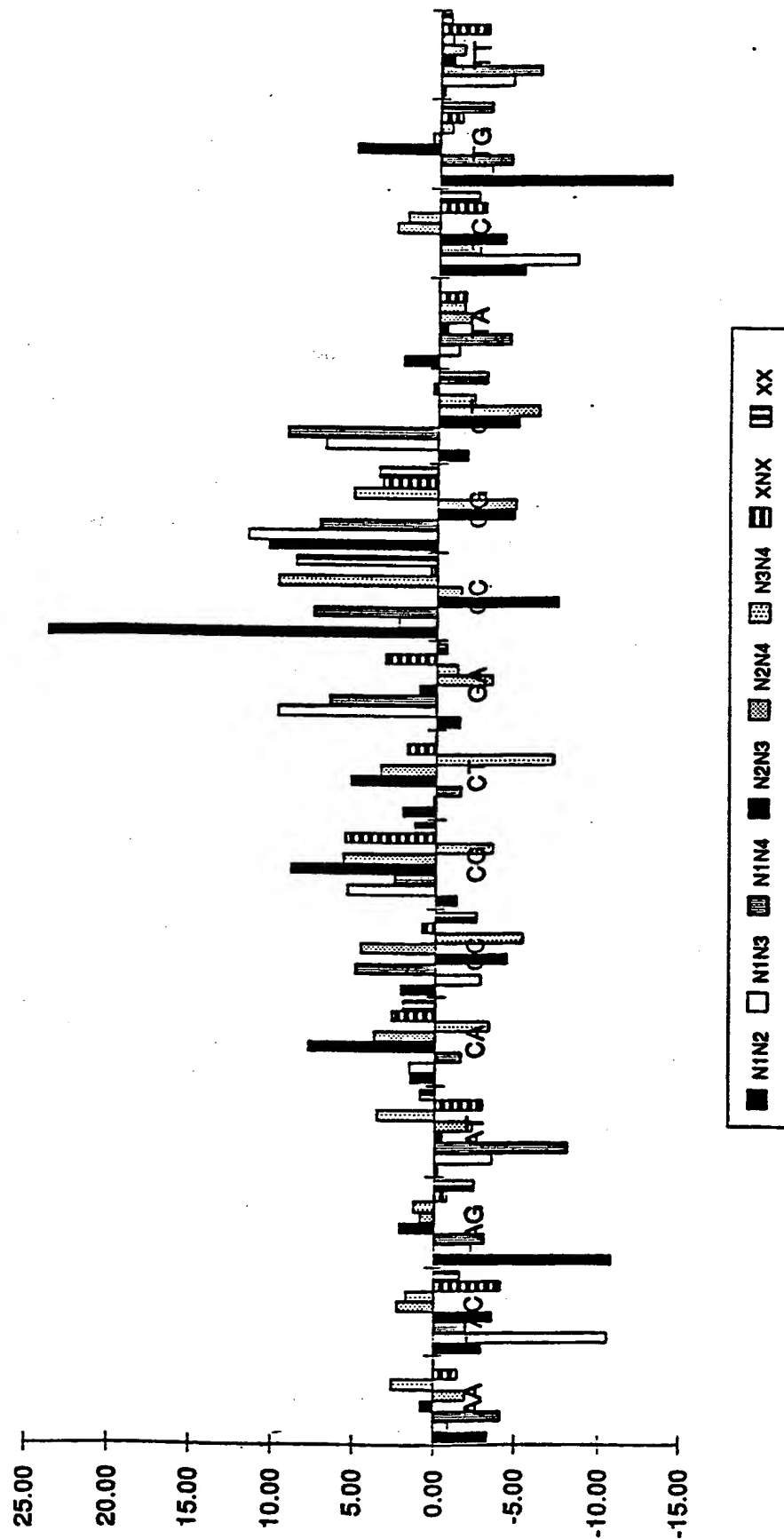


Fig 23

[illegible]

Fig. 24

35 / 39

Fig. 25

TTCCTTCC	TTACTTCC	TACCTTCC	TAACTTCC
TTCCTTAC	TTACTTAC	TACCTTAC	TAACTTAC
TTCCTACC	TTACTACC	TACCTACC	TAACTACC
TTCCTAAC	TTACTAAC	TACCTAAC	TAACTAAC

Fig. 26

TTCCNTTCC
TTCCNNTTCC
TTCCNNNTTCC
TTCCNNNNTTCC

36 / 39

Fig. 27A

---> --UL9--> --->
 5'-GCGTANXYZZCGTTTCGCACTTXYZZCTTCGTCCCAAT-3'
 3'-CGCATNYXQQGCAAGCGTGAAYXQQGAAGCAGGGTTA-5'

Score
high

Fig. 27B

--UL9-->
 5'-GCGTANQQXYCGTTTCGCACTTQQXYCTTCGTCCCAAT-3'
 3'-CGCATNZZYXGCAAGCGTGAZZZYXGAAGCAGGGTTA-5'
 <--- <---

Score
low

Fig. 27C

---> --->
 5'-GCGTANXYZZTTTCACGCTTGCCXYZZCTTCGTCCCAAT-3'
 3'-CGCATNYXQQAAGTGCGAACGYXQQGAAGCAGGGTTA-5'
 <---UL9---

Fig. 27D

5'-GCGTANQQXYTTTCACGCTTGCCQXYCTTCGTCCCAAT-3'
 3'-CGCATNZZYXAAGTGCGAACGZZYXGAAGCAGGGTTA-5'
 <--- <---UL9--- <---

Fig. 27E

---> --UL9-->
 5'-GCGTANXYZZCGTTTCGCACTTQQXYCTTCGTCCCAAT-3'
 3'-CGCATNYXQQGCAAGCGTGAZZZYXGAAGCAGGGTTA-5'
 <---

Fig. 27F

--UL9--> --->
 5'-GCGTANQQXYCGTTTCGCACTTXYZZCTTCGTCCCAAT-3'
 3'-CGCATNZZYXGCAAGCGTGAAYXQQGAAGCAGGGTTA-5'
 <---

Fig. 27G

--->
 5'-GCGTANXYZZTTTCACGCTTGCCQXYCTTCGTCCCAAT-3'
 3'-CGCATNYXQQAAGTGCGAACGZZYXGAAGCAGGGTTA-5'
 <---UL9--- <---

Fig. 27H

--->
 5'-GCGTANQQXYTTTCACGCTTGCCXYZZCTTCGTCCCAAT-3'
 3'-CGCATNZZYXAAGTGCGAACGYXQQGAAGCAGGGTTA-5'
 <--- <---UL9---

37 / 39

HIVBH101 (HIV LTR sequence)
 GTTAGAGTGG AGGTTTGACA GCCGCCTAGC ATTCATCAC ATGGCCCGAG
 AGCTGCATCC GGAGTACTTC AAGAACTGCT GACATCGAGC TTGCTACAAG
 <<NF-κB>>| <<<NF-κB>>| <Sp-1 III| <Sp-1 II| <
 GGACTTTCCG CTGGGGACTT TCCAGGGAGG CGTGGCCTGG GCGGGACTGG
 Sp-1 I>| TATA|
 GGAGTGGCGA GCCCTCAGAT CCTGCATATA AGCAGCTGCT TTTTGCCTGT
 +1 prim transcript start --->
 ACTG GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTC

Fig. 28

Fig. 29A

where X is the number of bases in the test site.

Fig. 29B

5'-GCAGAATTCTGCAGNNNCGTTCGCACTTTCTAGAGCTCAGG-3'

Fig. 29C

5' -GCAGAATTCTGCAGNNNNNNNNCGTTCGCACTTTCTAGAGCTCAGG-3'

Fig. 29D

5'-GCAGAATTCTGCAGCGTTCGCACTTNNNNNNNNNTCTAGAGCTCAGG-3'

39/39

UL9 Site 3' relative to the test sequence:

primers
EcoRI/PstI
 5'-CGTGAATTCTGCAG-3'
 5'-CGTGAATTCTGCAGATG-3'

Asp718/RsaI/KpnI
 restriction

site UL9 site
 5'-CGTGAATTCTGCAGATGAGGTACNNNNNNCGTTCGCACTTTCTAGAGCTCTCC
 test
 site

GTGAAAGATCTCGAGAGG- 5'
 AAGATCTCGAGAGG- 5'
XbaI/SacI
primers

Fig. 30

Small Molecule Binding Sequence	Expected Score in Assay	Potential Test Site Sequence
<u>UL9 site</u> 5'-...CGTTCGCACTTTTAC...-3'	high	TTAC
5'-...CGTTCGCACTTTACN...-3'	high	TACN
5'-...CGTTCGCACTTACNN...-3'	high	ACNN

Fig. 31

Small Molecule Binding Sequence	Expected Score in Assay	Potential Test Site Sequence
<u>SmaI</u> 5'-...CCCGGGTTAC...-3'	high	TTAC
5'-...CCCGGGTACN...-3'	low	TACN
5'-...CCCGGGACNN...-3'	low	ACNN

Fig. 32

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/12388

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : Please See Extra Sheet.

US CL : 435/6, 235; 514/44; 530/350, 351; 536/23.1, 23.2

According to International Patent Classification (IPC) r to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 235; 514/44; 530/350, 351; 536/23.1, 23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CAS ONLINE, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 5,071,773 (EVANS ET AL) 10 December 1991.	1-44
A	US, A, 5,096,815 (LADNER ET AL) 17 March 1992.	1-44



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be part of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 24 MARCH 1994	Date of mailing of the international search report APR 19 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile N . (703) 305-3230	Authorized officer GIAN WANG <i>Jee Warden for</i> Telephone N . (703)-308-0196

INTERNATIONAL SEARCH REPORT

national application No.
PCT/US93/12388

A. CLASSIFICATION OF SUBJECT MATTER: IPC (5):

C12Q 1/68; C12N 7/00; A01N 43/04; A61K 31/70; C07K 3/00, 13/00, 15/00, 17/00; C07H 17/00